

Technical Report 1050

Training Critical Thinking Skills for Battlefield Situation Assessment: An Experimental Test

Marvin S. Cohen and Jared T. Freeman
Cognitive Technologies, Inc.

Jon J. Fallesen
U.S. Army Research Institute

F. Freeman Marvin and Terry A. Bresnick
Cognitive Technologies, Inc.

October 1996

19970207 003

DTIC QUALITY INSPECTED 4



**United States Army Research Institute
for the Behavioral and Social Sciences**

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel**

**EDGAR M. JOHNSON
Director**

Research accomplished under contract
for the Department of the Army

Cognitive Technologies, Inc.

Technical review by

James W. Lussier
Michael J. Singer

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to : U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-STP, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1996, October		2. REPORT TYPE Final		3. DATES COVERED (from... to) January 1994-September 1996	
4. TITLE AND SUBTITLE Training Critical Thinking Skills for Battlefield Situation Assessment: An Experimental Test				5a. CONTRACT OR GRANT NUMBER MDA903-92-C-0053	
				5b. PROGRAM ELEMENT NUMBER 0602785A	
6. AUTHOR(S) Marvin S. Cohen (CTI), Jared T. Freeman (CTI), Jon J. Fallesen (ARI), F. Freeman Marvin (CTI), and Terry A. Bresnick (CTI)				5c. PROJECT NUMBER A790	
				5d. TASK NUMBER 1111	
				5e. WORK UNIT NUMBER C04	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Cognitive Technologies, Inc. 4200 Lorcom Lane Arlington, VA 22207				8. PERFORMING ORGANIZATION REPORT NUMBER 95-5	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-PK 5001 Eisenhower Avenue Alexandria, VA 22333-5600				10. MONITOR ACRONYM ARI	
				11. MONITOR REPORT NUMBER Technical Report 1050	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES For related training materials see ARI Research Product, Methods for Training Cognitive Skills in Battlefield Situation Assessment by Jared T. Freeman and Marvin S. Cohen COR: Jon J. Fallesen					
14. ABSTRACT (Maximum 200 words): In battlefield situation assessment, officers must interpret information that is incomplete, unreliable, and often conflicting and gather new information to improve their assessments and plans. In previous work, a framework for these cognitive activities was developed based on interviews with activity-duty command staff, and a training method was developed. That training helped officers to find and assess the reliability of hidden assumptions and to resolve conflicting evidence. Forty-three U.S. Army officers participated in an experimental training study with scenario-based tests. Trained officers generated more accurate arguments concerning a given assessment than did controls. Improvements in quality were related to the increased relevance of their judgments. In some problems, training countered a tendency to change hypotheses too readily; in other problems, training countered a tendency to hold on to a hypothesis too long. Training did not decrease confidence in evaluations, nor did it hypersensitize officers to information.					
15. SUBJECT TERMS Situation assessment Decision making Battle Command Training Critical thinking Planning Cognition Metacognition					
SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	63	

Technical Report 1050

Training Critical Thinking Skills for Battlefield Situation Assessment: An Experimental Test

Marvin S. Cohen and Jared T. Freeman

Cognitive Technologies, Inc.

Jon J. Fallesen

U.S. Army Research Institute

F. Freeman Marvin and Terry A. Bresnick

Cognitive Technologies, Inc.

Fort Leavenworth Research Unit

Stanley M. Halpin, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

October 1996

**Army Project Number
2O262785A790**

**Personnel Systems and
Performance Technology**

Approved for public release; distribution is unlimited.

FOREWORD

The U.S. Army recognizes that assessment of battlefield situations is critical to the tactical and operational success of its fighting organizations. Yet conventional training does not treat situation assessment as an explicit skill. Instead it is treated as incidental to the teaching of tactics and experience in tactical planning. What the conventional approach misses is the attention to an individual's thought process. An experimental program was conducted as part of a larger program of research by the U.S. Army Research Institute for the Behavioral and Social Sciences' Fort Leavenworth Research Unit to test the application of cognitive psychology to the improvement of battle command abilities. The program of cognitive research was conducted at the request of the Commander of the U.S. Army Training and Doctrine Command in 1994.

The research documented in this report concerns the testing of a program of instruction for improving critical thinking skills. The instruction focuses on safeguarding against uncertain or unreliable information and handling conflicts in the information. This study is unique not only because it approached officer instruction from a cognitive skills perspective, but also because it provides empirical evidence that such an approach has merit for improving thinking skills for battle command and decision making.

ZITA M. SIMUTIS
Deputy Director
(Science and Technology)

EDGAR M. JOHNSON
Director

ACKNOWLEDGMENTS

We wish to acknowledge and thank Mr. Rex Michel of the U.S. Army Research Institute for the Behavioral and Social Sciences, who was the author of the Arisle Island scenario, upon which the test questions were based. We are also indebted to Major Clay Miller, who obtained and scheduled participants for this study. Finally, thanks to General (U.S. Army retired) Charles P. Otstott for his valuable advice and assistance.

TRAINING CRITICAL THINKING SKILLS FOR BATTLEFIELD SITUATION ASSESSMENT: AN EXPERIMENTAL TEST

EXECUTIVE SUMMARY

Research Requirement:

For at least a decade battlefield situation assessment has been openly recognized by the U.S. Army as a key component to tactical decision making. The recent emergence of naturalistic models of decision making has also underlined the importance of identifying how people actually think and decide, especially how they interpret what the real problems are and what to do about them. When these two influences are combined, it is easy to recognize that conventional training has not sufficiently addressed the actual ways commanders and staffs assess situations. This research was aimed at exploring whether the application of a cognitive skills approach to training could improve battlefield situation assessment.

Procedure:

A cognitive framework was developed and documented in earlier research. This framework was referred to as the recognition/metarecognition (R/M) model (Cohen, Adelman, Tolcott, Bresnick, & Marvin, 1994). Using this model, midgrade Army officers were examined while they conducted battlefield planning. An interesting tendency was identified: Proficient decision makers appear to construct complete and coherent situation models by collecting or retrieving information and resolving any apparent conflicts. The decision maker's focus on a situation model indicates what needs to be believed if the model is to be accepted. An individual's evaluation of a situation model can be reduced to the testing of the reliability of the underlying assumptions. Instruction was developed to focus on this particular skill. Specifically, the instruction was designed to help officers identify assumptions hidden in their assessments.

The experiment to test the merit of the instruction compared pretest and posttest scores between officers who received the training and those who did not. Twenty-nine officers received the training and 8 officers served as controls. Training took 90 minutes. The first part consisted of having the individuals reflect on a personal experience where they were completely confident of an assessment and showing them how and why that "certainty" could be questioned. The training, referred to as the crystal ball technique, forces the officers to come up with alternative assessments. By doing this, assumptions are exposed that were hidden in the original understanding of the situation. In the second part of the training, officers are asked to reflect on personal experiences when they were surprised. Instead of disregarding new, conflicting cues, the training shows how to reinterpret the new information or to create a new situation model that accounts for all, previously conflicting information. Tactical examples are used to demonstrate and practice the techniques.

The materials used for testing involved a scenario about the invasion of an island nation and the U.S. response. The task used to measure performance focused on an officer's ability to evaluate a specified assessment of a situation. Officers rated their agreement to an assessment, read updated information, and rerated their agreement to the original assessment. They explained each rating in writing. The measures included the number of reasons used to explain a rating, the quality of these reasons, the number of relevant arguments given, and accuracy of the ratings compared to those from subject matter experts. Data were analyzed by separate problems. The effects of training were also analyzed to see if confidence in one's own assessment was undermined or whether the training had oversensitized the officers to new information.

Findings:

The training on critical thinking skills helped officers generate more accurate arguments without decreasing confidence in their explanations or hypersensitizing them to new information. Improvements in accuracy were probably related to an increase in relevant explanations that supported or opposed plausibility judgments of the assessments.

The training also appeared to counteract possible decision biases. The training tempered disconfirmation and confirmation bias for selected problems, not by discouraging the bias but by displacing the possible biases with critical thinking strategies. The trained officers also endorsed the training. They reported positive impressions and felt that the techniques would be useful in the field and should be integrated into formal Army courses.

Utilization of Findings:

The training techniques were adapted for use in an experimental battle command class for the Command and General Staff Officer Course (CGSOC). This application of the training focused on finding hidden assumptions. Students in that class felt that the instruction led to a gain of 20 percent in their expertise. This and other cognitive skill techniques have been incorporated into a new required core course for CGSOC on critical thinking.

The findings also indicate the merit in an approach that uses a cognitive skill focus to enhance battle command performance. The R/M model suggests other cognitive skills that could be targeted for improvement. For example, the "quick test" serves a metacognitive function to indicate when there is sufficient reason and time to enter into critical assessment. Another aspect that was not investigated in the present research was the use of the form of the situation models and knowledge structures relating to situation assessment. Instruction based on prototypical models (or story structures) might help the individual distinguish between what is typical and what is surprising in a specific situation. Such discriminations can help in the assessment of plausibility. Inclusion of additional components to the training will require additional time and development, but would also allow more proactive application of the skills to more challenging and complex situations.

TRAINING CRITICAL THINKING SKILLS FOR BATTLEFIELD SITUATION ASSESSMENT: AN EXPERIMENTAL TEST

CONTENTS

	Page
Introduction	1
Recognition/Metacognition Model	2
An Implementation of R/M Training	5
Hypotheses and Research Questions	8
Method	12
Design	12
Participants	12
Materials	12
Procedure	14
Analysis	15
Tests of Hypotheses	15
Exploration of Research Questions	26
Subjective Evaluation of the Training	34
Conclusions	37
References	41
Appendix A: Abbreviated Arisle Scenario	43
Appendix B: Problem Statements	49

LIST OF FIGURES

Figure 1.	Ways in which correcting steps can expose new problems, which are addressed in subsequent cycles of critiquing and correcting.	5
2.	There was no statistical difference between trained officers and the control group on the number of arguments generated.	17
3.	The largest number of posttest arguments was generated by trained officers who produced the most pretest arguments.	18
4.	Trained officers appeared to generate more arguments than controls on both problem parts but the difference was not significant.	19
5.	There was no significant difference between trained officers and the control group on the overall quality of arguments generated during the posttest.	20

CONTENTS (Continued)

	Page
Figure 6	
Trained officers generated significantly fewer neutral statements after training.....	21
7. The graphs present the proportion of arguments of each type generated by participants in both training conditions on both problem parts.....	21
8. Mean agreement ratings were lower for trained officers than controls on problem 5, lower on part B than on part A, and differed more between groups on part B	23
9. Ratings by trained officers were higher than those by controls, and thus closer to the SMEs' rating of 5, on both parts of problem 12.....	24
10. On problem 5, trained participants and controls differed more on problem part B than problem part A.....	25
11. Officers who initially agreed with the assessment for problems 5, 9 and 12 interpreted new information as disconfirming.....	28
12. Training appeared to moderate the disconfirmation bias on problems 9 and 12	29
13. Training appeared to moderate the confirmation bias on problems 2 and 10	30
14. Confidence was lower among trained officers than controls on part A of problem 5 but significantly higher on part B.....	32
15. The most frequent participants' ratings of the training were overwhelmingly positive.....	35

TRAINING CRITICAL THINKING SKILLS FOR BATTLEFIELD SITUATION ASSESSMENT: AN EXPERIMENTAL TEST

Introduction

Two young officers fly their helicopter on a border patrol mission over the snow-covered terrain. The navigator works by map, stopwatch, and compass. He plots their path on the map from one outpost symbol to the next, through one set of concentric elevation marks to another. The landscape conforms well enough to the map. Neither officer can spot some of the bunkers they expect to see on the ground, but they reason that the snow obscures the small outposts. The configuration of hills sometimes defies their expectations, but perhaps the shadows thrown by the setting sun are distorting their vision. At any rate, many of the landmarks they seek are in place, and that is an adequate confirmation of their course. Suddenly, their main rotor engine begins to cough. They descend to inspect it in friendly territory. As they touch down, an enemy rocket explodes under the helicopter, killing one officer and wounding the other. The officers were not on friendly ground, but lost in enemy territory.

An Army soldier sits guard duty at the periphery of his camp. It is night, but preparations for tomorrow's battle are underway. He listens to the sounds of his unit's light equipment moving into place, and the occasional noises of his fellow troops preparing their gear. As the sun rises, his camp is overrun by enemy forces, who moved into position protected by darkness...and by the guard's mistaken assumptions.

The division commander of a U.S. contingency force is defending a port through which reinforcements are arriving. Enemy forces to the north are commanded by an experienced officer, highly skilled, and well-equipped to cross the rivers in his path to the port. However, his heavy armor must traverse poor roads. To the south is a second enemy contingent. Its commander is less able, and he has fewer bridging assets, but roads are excellent and lead directly to the southern port. Furthermore, the southern commander has been successful in several recent advances towards the port. The U.S. commander reasons that the enemy's Soviet doctrine, to exploit success, will lead him to attack the port from the south, while perhaps conducting a diversionary engagement in the north. Then, events unfold that seem to contradict this assessment. Some of the enemy's southern forces are observed moving into his northern territory; the southern enemy destroys a bridge in the path of his own advance; the enemy initiates radio silence in the south, and then in the north as well. Do these events disconfirm the assessment of the U.S. commander. Are they signs of subterfuge that support it? How should he interpret the evidence before him?

In these scenarios, Army officers seek to assess complex and ambiguous circumstances. In no domain is this a simple task. In few is it as critical as in Army battlefield planning and operations.

The most popular paradigms for analyzing situation assessment, and decision making generally, fall into two categories: analytical and recognition. The analytical approach (e.g., Keeney and Raiffa, 1976) prescribes a form of decision making that attempts to be highly rational, one in which expertise resides in the ability to identify potential options and their outcomes, evaluate the utility of the outcomes along various dimensions, assess the probabilities

of the outcomes, and then to computationally combine the probabilities and utilities in order to compare the options. In contrast, the simplest recognitional approach describes expert decision making in terms of perceiving events, recognizing them to fit some known pattern, and responding with a familiar label or plan of action.

Both analysis and recognition are appropriate models of decision making in some contexts, but not in all. Battlefield events are often so complex and interdependent that they cannot be modeled from discrete components with known parameters, nor is there necessarily time to do so. The analytic approach is inappropriate in these situations, and empirical studies indicate that it is very rarely used. Evidence for recognitional decision making is ample and seems to increase with the experience of the decision maker (e.g., Larkin, McDermott, Simon, & Simon, 1980). But recognition fails to account for successful decision making in the face of novel events, that is, in situations that are at least in part unrecognizable. In highly complex battlefield situations, novelty is the rule rather than the exception. Moreover, it is in the enemy's interest to conceal the true pattern of his actions, and to utilize misleading patterns for purposes of deception. To cope with such situations, it is necessary to go beyond recognition. Experienced decision makers develop strategies for testing the validity of recognitional responses and for controlling recognitional processes and modifying their results. In short, experienced officers are capable of perceiving when recognition is weak, critiquing their assessments, and improving them.

We have argued that these strategies can be regarded as *meta-recognitional*, by analogy to other executive, or metacognitive, strategies that monitor and regulate more basic processes, such as meta-memory, meta-comprehension, and meta-attention (Gavelek and Raphael, 1985; Gordon and Braun, 1985; Kuhn, Amsel and O'Loughlin, 1988). We have developed a model of these decision-making process, called the Recognition/Metacognition (R/M) model. That model has been described more fully in a previous report (Cohen, Adelman, Tolcott, Bresnick, & Marvin, 1993). Training methods based on that model were described in detail in another previous report (Freeman & Cohen, in preparation). This report briefly recapitulates the R/M model (in the following section) and summarizes the training methods based on the R/M model (in the subsequent section). The remaining sections report the results of testing the training methods with active-duty Army officers. Testing materials are reproduced in the Appendices.

Recognition/Metacognition Model

Meta-recognition is a cluster of skills that support and go beyond the recognitional processes in situation assessment. Situation assessment begins as recognition but continues if there is cause and opportunity to do so with one or more cycles of critical thinking. In a process called *critiquing*, the decision maker looks for sources of uncertainty, such as: 1) incomplete information; 2) unreliable assumptions; or 3) data that support conflicting conclusions. When problems are found, they are the targets of a *correction* process, in which the decision maker collects more information, retrieves more information from long-term memory, or adjusts assumptions that stand in for missing information. The decision maker's newly elaborated understanding of the problem is re-evaluated as situation assessment continues in further cycles of recognition and metacognition. Critiquing and correcting are regulated by a process called the *Quick Test*. They continue only as long as time is available, the cost of an error is high, and the situation remains unfamiliar or problematic.

The following example is based on think-aloud problem solving sessions with active duty

Army officers who were presented with a battlefield scenario. A division plans officer is trying to predict the location of an enemy attack. The enemy has had the greatest success in the south, which the enemy is likely to want to exploit; its most likely goal, city Y, is in the south; it has the best supplies in the south; and the best roads are in the south. The planner concludes that the attack will be in the south.

The normal, recognitional meaning of each cue (prior success, a lucrative goal, supplies, and roads) is to expect attack in the sector associated with the cue. If time is limited or the consequences of being wrong about the location of attack are not great, the planner will not consider the issue further. However, when the stakes are high, time is available, and the situation is not completely routine, he may not be content with this initial recognitional response.

Critiquing can result in the discovery of three kinds of problems with an assessment: *incompleteness*, *unreliability*, or *conflict*. An assessment is incomplete if key elements of a situation model or plan based on the assessment are missing. In identifying incompleteness, the recognitional meanings of the cues are embedded within a structure of some kind. In particular, story structures depict causal and intentional relations among events and have characteristic sets of components (Pennington & Hastie, 1993). In particular, the main components of stories concerned with assessments of enemy intent are *goals*, *capabilities*, and *opportunities* (which elicit) the *intent* to attack at a particular place and time (which leads to) *actions* (which result in) consequences. For example, an officer might conclude that the enemy's intent to attack in the south was adopted because of higher-level goals such as capturing city Y and exploiting prior success in the south, superior capabilities in the south by virtue of better supplies, and superior opportunity via better roads. Future actions that would be expected include removing obstacles in the relevant sector, massing artillery, and moving up troops.

In our example, the officer looks for an argument supporting the conclusion that the enemy will attack in the south based on *each component* of the story structure. He finds the story to be incomplete because none of the enemy actions expected to occur prior to an attack have yet been observed. More subtly, the story may also be incomplete because the officer has not fully considered the factor of capability. What about the relative strength of artillery, armor, and leadership in the north versus the south? Moreover, he has not fully considered the factor of accessibility. What about mountain or river crossings required in a southern versus a northern attack? Correcting steps may generate the information required to complete this story by directing the retrieval of prior knowledge, the collection of new observations or analyses, or the revision of assumptions.

Another function of critiquing is to find conflict, new arguments whose conclusions contradict the conclusions of existing arguments. In our example, the officer's further consideration of enemy capabilities produced an assessment that both troop strength and leadership were superior in the north. The normal, recognitional meanings of these assessments are that the enemy intends to attack in the north. Moreover, fleshing out the accessibility component of the story produced another conflicting argument: The northern forces had superior river crossing skills, making the northern route easier on the whole.

Critiquing can also expose unreliability in a situation model or plan. Understanding and planning is unreliable if the argument from evidence to conclusion, or from goals to action, is conditioned on doubtful assumptions. For example, taken by itself, troop movement toward the south is an unreliable indicator of attack in the south since there may be even more troops

moving north, or the enemy may intend to move the observed troops north at the last minute.

Unreliability is different from conflict, because here critiquing neutralizes the argument for attack in the south based on troop movements but does not provide an argument *against* attack in the south.

Critiquing and correcting for one problem may lead to the creation and detection of other problems. In this example, efforts to create a complete story led to discovery of the conflict between better capabilities and accessibility in the north versus more plausible goals in the south. The officer resolved this conflict by rejecting the normal, recognitional meaning of the evidence favoring attack in the north. He generated an alternative interpretation of these same data, that the *main* attack will be in the south but that a *diversionary* attack is planned for the north. This resolution of the conflict, however, opened the door to a new problem: unreliability of the assumption about a diversionary attack in the north.

Figure 1 summarizes how steps of critiquing and correcting can be linked in the R/M framework. The three types of problems explored by critiquing are shown as three points on a triangle, representing model incompleteness, unreliable assumptions in arguments for the key assessment (e.g., intent to attack in the south) or in rebuttals of arguments against the key assessment, and the existence of conflicting arguments that contradict the key assessment. The arrows showing transitions from one corner of the triangle to another represent correcting steps. It is these correcting steps that may sometimes, but not always, produce new problems. For example, correcting incompleteness in the situation model by retrieving or collecting data or by making assumptions can lead either to unreliable arguments or to conflict with other arguments. Resolving conflict by critiquing a conflicting argument can lead to unreliable assumptions in rebuttals. Dropping or replacing unreliable assumptions can restore the original problems of incompleteness or conflict. These new problems may then be detected and addressed in a subsequent iteration of critiquing.

Our analysis of 34 critical incident interviews with Army command staff suggests an important feature of naturalistic decision making related to Figure 1. Proficient decision makers first try to fill gaps and explain conflict, and only then assess the reliability of assumptions. Thus they tend to advance from the upper right and left corners of the triangle down to the bottom, converting problems of incompleteness and conflict into problems of unreliability. In short, they try to construct *complete and coherent situation models*. They do this if possible by means of newly collected or retrieved information, but if necessary by adopting assumptions. Success in filling gaps and resolving conflict does not mean that decision makers accept the resulting situation model. But it does tell them what they must believe *if they* were to accept it. This process facilitates evaluation of a model by reducing all considerations to a single common currency: the reliability of its assumptions. If unreliability is too great, a new cycle of critiquing will hopefully expose it and trigger efforts to construct a new story.

The R/M model describes a set of skills that supplement pattern recognition in novel situations. These skills include identifying key assessments and the recognitional support for them, checking stories and plans based on those assessments for completeness, noticing conflicts among the recognitional meanings of cues, elaborating stories to explain a conflicting cue rather than simply disregarding it, sensitivity to problems of unreliability in explaining away too much conflicting data, attempting to generate alternative coherent stories to account for data, and a sensitivity to available time, stakes, and novelty that regulates the use of these techniques.

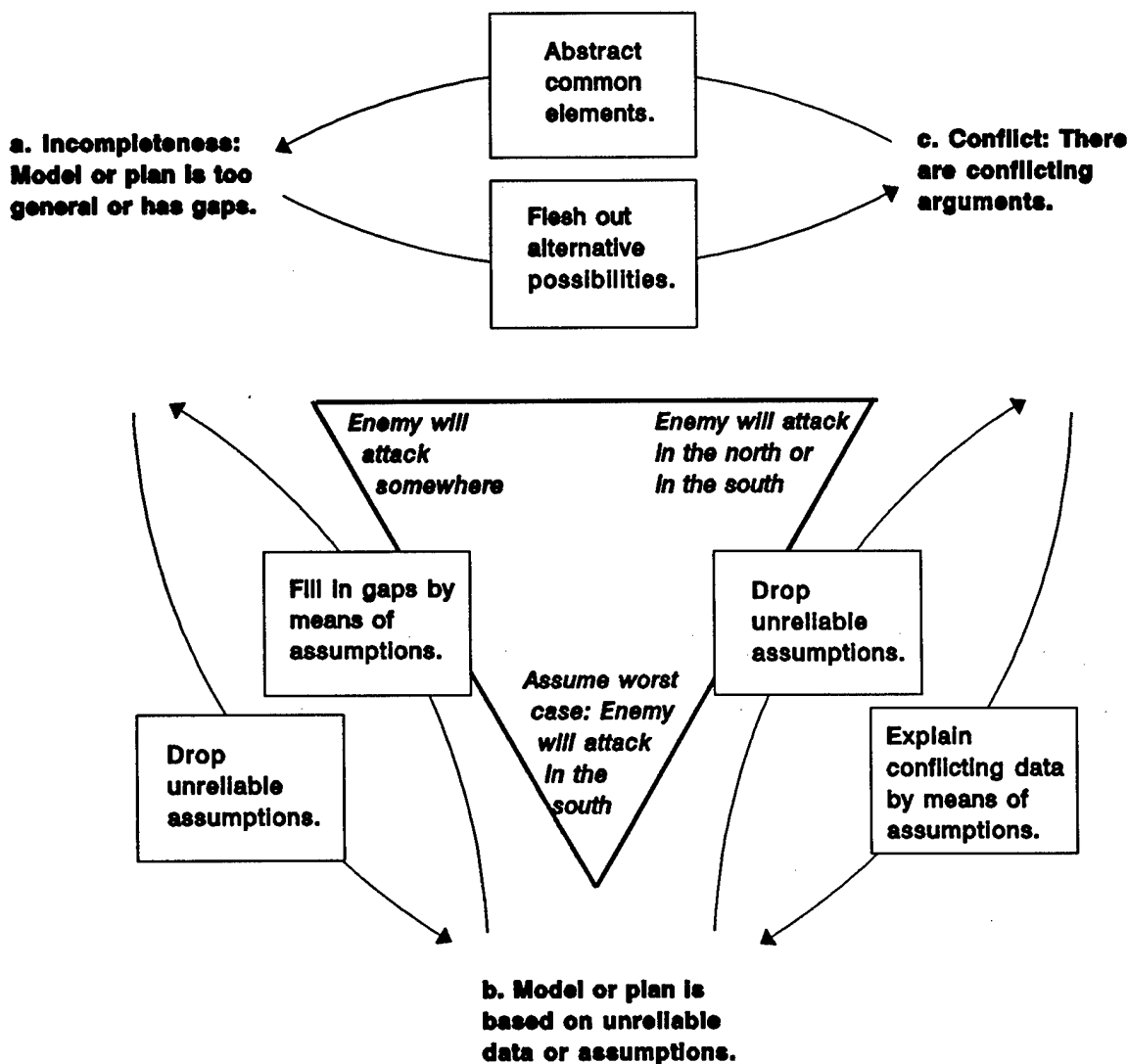


Figure 1. Ways in which correcting steps can expose new problems, which are addressed in subsequent cycles of critiquing and correcting.

These skills are neither as domain-specific as simple pattern recognition, nor as general-purpose as analytical methods. Like analytical tools, meta-recognitional skills may be applicable with minor adaptations across a wide range of domains. Unlike analytical skills, however, their use requires a relatively strong base of familiarity in a domain. They build upon the knowledge embedded in recognitional skills, but do not by any means replace it.

An Implementation of R/M Training

We have prepared a training program for meta-recognition skills in Army command staff battlefield situation assessment. It is designed to hone the situation assessment performance of U.S. Army officers by improving their critiquing and correcting skills. In particular, it is intended to help officers identify assumptions hidden in their assessments, to explain anomalous events,

to evaluate the plausibility of these explanations, and to generate new assessments when necessary. In the training sessions we have conducted, officers read a handbook containing explicit instructions about critiquing assessments along with numerous examples, listen to brief, summarizing lectures, and participate in individual and group exercises based on realistic military scenarios and their own experiences. In the following paragraphs, we describe the training. Training materials may be found in Freeman & Cohen (in preparation).

The Army training has two major segments. One focuses on situations in which the decision maker feels relatively certain of his or her conclusions, and focuses on critiquing and correcting for unreliability. The other focuses on detecting and handling conflicting observations. We will very briefly convey some of the flavor of each of the two Army segments.

Handling "Certainty"

We begin the discussion by asking officers for a personal experience in which they felt completely certain of some assessment. We then show how that "certainty" could be questioned. This method forces the officers to generate an alternative story that covers all the evidence. In doing so, it exposes assumptions underlying the story that they currently accept, and helps them evaluate the story for reliability. These assumptions can be evaluated and, if time and stakes warrant, can be checked. Appropriate correcting steps can be taken when weaknesses in the story are found. In the end, even if officers retain the original story, their confidence in it will have been earned.

The following example of "certainty" was volunteered in one class. A battalion officer, facing an enemy across the river, predicted that they would cross the river at point X. Point X was relatively close to the enemy's present position, the river at point X was relatively shallow, and a combination of vegetation and terrain there would provide concealment. He recommended concentration of friendly forces in the vicinity of point X.

The crystal ball method for finding hidden assumptions consists of four steps:

1. Select a critical assessment, no matter how confident you are that it is true (e.g., that the enemy will cross the river at point X).
2. Imagine that a perfect intelligence source, such as a crystal ball, tells you that this assessment is wrong.
3. Explain how this assessment could be wrong.
4. The crystal ball now tells you that your explanation is wrong and sends you back to step 3. (Continue until the set of exceptions to your original conclusion seems thorough and representative and the ways it could go wrong.)

After each new exception was mentioned, the crystal ball told the trainee, "No, that's not the reason why the assessment is wrong. Come up with another explanation." In this particular case, the crystal ball method elicited a number of ways this "certain" assessment might fail: (i) The enemy might anticipate that our force will be at point X and decide not to cross there. (ii) The enemy might detect the movement of our force to point X and decide not to cross there. (iii) There are good crossing sites that we missed. (iv) The enemy doesn't know how good a

location point X is. (v) The enemy doesn't have any river crossing assets. He can't cross the river at all. (vi) The enemy's river crossing assets are so good that he can cross elsewhere. (vii) The enemy has a large enough force that they can accept casualties in crossing elsewhere. (viii) The enemy's objectives are not what we thought. He doesn't need to cross the river. (ix) The enemy will use air assault forces to get across the river.

Usually, trainees are surprised at the quantity and the plausibility of the exceptions that the crystal ball method elicits from them. They now realize that the original assessment rested on the assumption that none of these exceptions was true. However, the existence of these possible exceptions is not adequate cause to abandon that assessment. The next step is to evaluate the exceptions. Each one should be considered, at least briefly. The class is asked how they would handle each one. Some possible exceptions may be implausible, for example, that the enemy can afford large casualties. Some can be tested by data collection or by requesting additional intelligence, for example, that the enemy has superior river crossing assets. Other exceptions may motivate a change in plans to make them less likely. For example, to avoid anticipation or detection of our forces at point X, we might position our forces elsewhere, then move to point X later. Other exceptions may cause adjustments in planning to handle them in case they turn out true. For example, we might place reserves on paths behind the river in case we missed some sites or the enemy missed point X. Exceptions may also cause the adoption of a contingency plan. For example, if the enemy's objective turns out to be on the other side of the river, we might prepare to cross the river ourselves. Finally, some exceptions might have to be accepted as known risks, for example, if the enemy uses air assault.

Handling Conflicting Data

In the second unit of training, we ask officers to describe personal experiences in which they were surprised; for example, the enemy attacked in an unexpected sector. We then ask if any cues or indicators had been observed that, in hindsight at least, could have served as a warning. Typically, such cues are clearly remembered, but were disregarded at the time.

A common response to observations that conflict with a previous conclusion is to disregard or discount them. Another response to conflict, which may be equally bad, is to lose confidence and immediately abandon the original assessment. An unexpected event means that situation understanding is imperfect, but the fault may not lie in the original assessment. It may lie in an incorrect interpretation of the new event. In situations where no patterns fits all the data, the correct assessment *must* involve some "explaining away" of conflicting data.

In this training segment, officers learn to monitor or critique for conflicting evidence and learn how to handle conflicting observations when they occur. When conflict is detected, they begin by modifying the current story to explain the surprising events in terms of their original assessment. They then evaluate the reliability of the resulting story. If the explanations prove to be implausible, officers alter the assessment itself and create a new story. The procedure consists of these steps:

1. Notice unexpected events.
2. Explain an unexpected event in terms of your current assessment. If there have been previous unexpected events, try to find the simplest reliable explanation covering all of them.
3. Evaluate the reliability of your account of all the unexpected events.

4. If the explanation is not reliable, change your assessment and return to step I.

The crystal ball technique can be useful in generating explanations of conflicting data. Now, however, the crystal ball tells you that the original assessment is *correct*, despite the conflicting observation, and asks you to explain how this could be. The crystal ball rejects each explanation the trainees generate and asks them to find another.

In one example, the enemy is expected to advance along a southern route but unexpectedly bombs a bridge in its own presumed path of advance. The crystal ball insists that the predicted advance by the enemy is correct despite the destruction of the bridge, and demands an explanation. The destruction of the southern bridge might be consistent with a main attack in the south, if: (i) The bridge was destroyed to prevent our troops from being reinforced; (ii) the enemy has better bridging equipment than we thought; (iii) destruction of the bridge was a mistake; (iv) it was part of a deception; or (v) the bridge was destroyed by our own troops rather than by the enemy.

A list of this kind may never be exhaustive. Nevertheless, it provides an understanding of the *kinds* of ways in which the current assessment could still be true despite a conflicting observation. To hold onto the assessment, it is not necessary to know which, if any, of these explanations is the case. But some such story elaboration must be true if the original assessment is to be maintained. Thus, the original assessment is no more reliable than the best of these explanations. These explanations may also point to ways that the assessment can be tested.

If there is more than one conflicting event, the officer must try to construct an overall story that most convincingly accounts for all the discrepant events. If more than one conflicting event can be explained in the same way, the story is more plausible. For example, two surprising events may have been reported by the same unreliable source, or they may represent the same enemy tactical plan. The fewer separate explanations, the less testing is required to verify the story, or - if testing is not possible - the fewer assumptions are required in order to hold onto the current assessment. However, these explanations must be individually plausible. Explaining away everything in terms of an enemy master plan for deception may be simple, but is not always convincing. The training proceeds to illustrate and discuss the dangers of explaining away too many conflicting cues.

To help officers take a new perspective on the problem, we advise them to disregard the prior assessment and instead to focus on the list of discrepant events that led them to abandon the assessment. They must answer the question, "What is the single most plausible explanation for these unexpected events?" This becomes the new assessment. The evidence supporting the prior assessment now becomes discrepant with respect to the new assessment, and must be explained by applying the procedures taught above. If a plausible account for these discrepant events can be found, the new assessment may be accepted. If not, the decision maker must generate another assessment and try to defend it.

Hypotheses and Research Questions

The training we have just described had several goals. First, it was designed to help Army officers generate and evaluate alternative assessments of a battlefield situation. Second, it was designed to help them notice data that conflicted with an assessment and to evaluate the impact of those data. Third, the training was intended to improve the accuracy of the

assessments that officers eventually settled on. We formulated three hypotheses to test the success of the training:

- Hypothesis 1. Trained participants will generate more and/or better arguments regarding the validity of assessments than will controls.
- Hypothesis 2. Trained participants will generate more arguments that conflict with an assessment than will controls.
- Hypothesis 3. Trained participants will evaluate assessments more accurately than controls, relative to subject matter experts.

In addition to these hypotheses, there were several research questions that we wished to address. These questions pertain to the potential influence of training on the direction of change in belief in response to new evidence, the magnitude of change in belief in response to new evidence, and the magnitude of confidence in beliefs.

Question 1. Various researchers have identified supposed biases in the interpretation of new evidence (Nisbett and Ross, 1980; Tversky and Kahneman, 1980). These researchers adopt relatively simple prescriptive models of inference, such as Bayesian updating, according to which a piece of evidence should have a fixed impact on belief regardless of the context of other evidence in which it occurs. (In Bayesian updating the impact of a piece of independent evidence on the conclusion is usually quantified as a likelihood function indicating relative support for different hypotheses.) Experimental research has shown that these models do not fit actual inference behavior. In particular, the weight that decision makers assign to a piece of evidence is not fixed; it may be influenced by their current belief regarding the conclusion. If new evidence disagrees with their current belief, it is more likely to be disregarded, discounted, or explained away. This effect is called *confirmation bias*.

The Recognition/Metacognition framework begins with very different premises. Unlike simple inference models, it does not regard the meaning of a piece of evidence as fixed. Rather, decision makers interpret evidence within the context of an on-going story-building process. The meaning of the same evidence may be quite different depending on the current story (which itself depends on the context of previous evidence). If evidence appears to conflict with the current story, assumptions may be required to make it fit. Explaining away conflicting evidence is thus part of the effort to make sense of data by constructing coherent situation models. However, skilled decision makers also step back and evaluate the stories that are built, and the assumptions required to build them. If too many implausible assumptions are required, they create, and then evaluate, alternative models.

Cohen (1993) has argued that "confirmation bias" behavior is appropriate under many circumstances. First, the attempt to generate explanations of discrepant evidence can shed light on the plausibility of a hypothesis. It exposes the assumptions that would have to be adopted if the hypothesis were to be accepted. This process provides a basis for evaluating the hypothesis, and often leads to testable predictions. Second, the process can lead to learning about cues that extends beyond the current situation. In the battlefield, the reliability of information is often unknown. An important indicator regarding the reliability of a piece of information is its degree of concordance with other information. If a large body of data supports one assessment, and there is a single outlier pointing to a different assessment, the discrepant information may not mean what it seems to mean. Information from that source may be scrutinized more carefully in

subsequent situations. Third, in many battlefield situations, no reasonable hypothesis fits the data perfectly; there is conflicting data no matter what assessment one adopts. Probabilistic combinations of possibilities (e.g., 55% chance of attack in the north, 45% chance of attack in the south) cannot be visualized, planned against, or acted on. Finding the most plausible explanation of discrepant evidence produces a single coherent picture of the situation which the decision maker can utilize (along with an understanding of its strengths and weaknesses) for planning and action.

The R/M training does not try to impose a fixed interpretation on cues. It focuses on generating explanations of conflicting data, and on exposing and evaluating the assumptions underlying such explanations. In some cases, this approach may lead to confirmation-bias-like behavior: Trained officers would be expected to explain away conflicting cues when the explanations are plausible (compared to the explanations that would be required to explain away the information *supporting* the hypothesis). In other cases, training might lead to disconfirmation-bias-like behavior. When the assumptions required to explain conflicting evidence become too numerous or too implausible, trained officers would be expected to reject the current story and generate another.

In the investigation of this research question, we will ask whether there is evidence for confirmation bias behavior and the impact of training on it. The analysis will rely heavily on the plausibility of various explanations, as reflected in a subject matter expert's assessments. We expect trained participants to be less likely than controls to disregard or explain away conflicting cues that should be taken seriously (i.e., when the explanations are not plausible), or to fail to explain away conflicting cues that are truly outliers (i.e., the explanations are plausible in the light of other evidence)

Question 2. Another way to think of the confirmation bias is as a sort of *primacy* effect, i.e., overweighting cues that occur early in a sequence and discounting cues that come later. *Recency* effects might also occur, i.e., overreacting to cues later in a sequence (and too hastily abandoning a hypothesis supported by early cues). These two effects would be reflected in the sensitivity of the participants to new information. Primacy effects would be reflected by smaller changes in belief after viewing new information; recency effects would be reflected by larger changes in belief after viewing new information. Our prediction was that training would not increase or decrease sensitivity across the board; trained officers would react more than controls to some new evidence (if it caused them to revisit and revise their assumptions about earlier evidence) and would react less than controls to other new evidence (if they found plausible explanations of the new evidence).

Question 3. Training might also have some effect on participants' confidence in their evaluations of given assessments. For example, if training made officers better able to evaluate information, it might increase their confidence. On the other hand, officers might be less confident if they found the training techniques to be confusing, or if the methods caused them to identify more conflict or to generate more interpretations of events than they could manage.

Question 4. Finally, we were interested in various potential sources of individual differences in performance. Specifically, did military tenure, rank, prior training, posts, or branch of Army service reliably account for skill in generating arguments concerning the plausibility of a given assessment?

In sum, we addressed the following research questions:

Question 1. What is the effect of training on the so-called confirmation bias?

Question 2. How does training affect the sensitivity of participants to new evidence?

Question 3. What is the effect of training on confidence in assessments?

Question 4. Which sources of individual difference predict ability to evaluate assessments?

To test our hypotheses and to explore these research questions, we performed a controlled study at two U.S. Army installations. In the remainder of this paper, we discuss the design, analysis, and conclusions of that study.

Method

Design

The study crossed two training conditions and two sequences of test materials in a pretest/posttest design. The training conditions were the instruction described in a previous section of this report, and a control condition, described below. The levels of problem order were two tests (discussed below) presented in a counterbalanced manner as pretest or posttest. Both variables, training and problem order, were between subjects factors in the design.

Participants

Forty-three U.S. Army officers participated in the present study. Data from 37 of these officers was used in the analyses, below¹. The 37 participants held the rank of First Lieutenant, Captain, Major, or Lieutenant Colonel, and were assigned to artillery (14 officers), aviation (2), engineering (2), infantry (15), or military intelligence (2). Two participants did not provide the branch to which they were assigned. Twenty nine of the officers received the experimental training treatment, at Fort Lewis. Eight officers at Fort Riley served as controls. Eighteen participants received one set of problems on the pretest and the other on the posttest; 19 participants received test materials in the reverse order. (See Appendix E).

The level of military experience among controls was high, at a mean of 13.25 years, relative to that of the treatment group, whose members had served for a mean of 10 years. Only one member of the control group had less than the overall median 10 years of military tenure. While the difference in years of military experience between groups was not statistically significant ($t_{35} = 1.66, p = .106$), it was an imbalance worth noting.

Officers were relatively well distributed between treatment groups by branch, though there were no controls who served in aviation or military intelligence. (See Appendix D).

Materials

The test materials consisted of a description of a military scenario and 12 problems regarding the scenario. Each problem consisted of two parts (A and B). Participants were asked to respond to two test questions on each part of each problem.

The scenario concerned the invasion of an island nation (Arisle) by its neighbor (Mainlandia) and the U.S. response. It consisted of a chronologically organized status report, a mission statement, a summary of U.S. forces, an intelligence estimate concerning the enemy's capabilities and situation, a description of Arisle, a detailed map of the island, and a large-scale map of Arisle and surrounding islands (see Appendix A).

¹Data from six of the 43 officers were dropped prior to analysis. Four officers in the training condition received only the written training materials, not the brief lectures and interactive exercises that were believed to be particularly helpful to later participants. One officer in the training condition arrived after the pretest had been completed. One control rushed to complete the posttest and depart for an appointment.

The problems each consisted of two parts, printed on separate pages (see Appendix B). The first part (designated part A) contained an assessment or conclusion regarding a particular topic, (e.g., the status of hostages held by enemy forces, the air defense capabilities of the enemy, the placement and mobility of enemy reinforcements), preceded by information regarding the relevant topic upon which the assessment was based. The second part of each problem (designated part B) consisted of the same information and assessment, plus new information. The new information was a mix of items that supported, disconfirmed, or were neutral with respect to the assessment. A typical example of a part A problem statement was the following:

The government of Arisle is a representative democracy with a governor and a five-member board of representatives elected by the people. The governor for the past nine years has been Quiton Pailou, who has created many reforms in education, taxation, and personal freedoms. He is greatly admired by the majority of the people, especially because the economy and standard of living has improved considerably during his administration. There is a generally cordial relationship between Pailou's government and the American Terrestria Corporation as well as the Japanese Pineapple Company.

ASSESSMENT: The great majority of the population does not support a Mainlandia take over. They can be expected to generally support a US invasion force.

In part B of this problem, the following was appended to the information and assessment, above.

NEW INFORMATION: Arisle was a possession of Mainlandia from the 12th to the 18th Century when it was captured by the French during the Napoleonic Era. It remained a French territorial possession until 1947, when it gained its independence. Recent rallies of the radical Arisle Revolutionary Front (ARF) political party have brought out large crowds with their message of "Arisle First!" A suspicious fire at the Japanese pineapple plantation last month is rumored to be the work of the ARF. Since the invasion by Mainlandia, no acts of defiance by the civilian population have been reported.

Participants who received training solved six of these two-part problems on the pretest and six on the posttest. Controls received the eight problems most frequently answered by trained participants, four per test. For each problem statement, participants were asked to respond in writing to two questions:

1. Please evaluate the assessment. In what ways is the reasoning good? bad?
2. Do you agree with the assessment? Use this scale:
 1. Strongly disagree
 2. Moderately disagree
 3. Don't know
 4. Moderately agree
 5. Strongly agree

The two problem sets appeared roughly equivalent in how subject matter experts

(SME's) responded to the second question. The experts were a retired U.S. Army Lieutenant General with 32 years of military experience and a Major in the U.S. Army reserves with 18 years of military experience. First, the problem sets were compared in terms of the plausibility of the assessments they contained. The average of the SMEs' ratings of agreement with the assessment on each problem were compared between problem sets. Ratings on the given five-point scale averaged 2.96 for one problem set, 3.25 for the other. The difference in SME's ratings between tests was not significant ($t_{22} = -.559, p = .583$). Second, the degree of consensus between the two SME's was compared between problem sets. The average absolute difference between the SME's in agreement was 1.25 for one problem set, and on the other it was 1.17. This difference between tests was not significant ($t_{22} = -.196, p = .847$). Interrater reliability for the agreement ratings by the two SME's over both problem sets was modest but acceptable (Pearson's $r = .536, p = .007$).

Procedure

Officers participated in the study in groups of three to six for approximately three and one-half hours, including breaks. Each session began with a brief introduction to the study. Participants then filled out a biographical survey form while researchers distributed pretest materials. Participants were asked to read background materials concerning Arisle for approximately 15 minutes. They then turned to the problem statements. For each part (A and B) of each problem, they wrote answers to the two test questions. Testing took approximately 40 minutes.

Experimental participants then received the training described above over a 90-minute period. Control participants generated their own assessments concerning several military scenarios and performed a psychological battery over the same time period. The psychological tests were selected largely to lend face validity to a supposed study of individual differences in situation assessment. They concerned spatial memory, learning style, need for cognition, intolerance of ambiguity, perfectionism, locus of control, and a test of several dimensions of behavior (friendly vs. unfriendly, dominant vs. submissive, and emotionally expressive vs. instrumentally controlled).

A 40-minute posttest followed training or control activities. Experimental participants concluded the session by completing a debriefing form.

Analysis

Tests of Hypotheses

Analytic Strategy

The primary interest in each analysis was the effect of treatment (trained vs. control) on a dependent variable of interest, such as the number of arguments generated when evaluating an assessment. However, we attempted to control other sources of variance in addition to treatment condition. These were prior skill as measured on the pretest, and, in some analyses, the order of the problem sets used as pretest and posttest, and/or problem part (A or B).

We employed a conditional, two-stage hypothesis testing strategy (see, for example, Pedhazur and Schmelkin, 1991). In the first stage of analysis, we tested for interactions of pretest score with the treatment and other independent variables. This stage of analysis employed a full multivariate regression model, including the treatment (training vs. control), pretest score, and other independent variables, plus all of their interactions. This analysis served as the basis for decisions regarding the possible simplification of the model to be used in the second stage. The decision rules for the analysis were as follows:

1. If pretest score interacted with any of the other independent variables at a confidence level of $p < .25$, then the full regression model was used.
2. If no interactions involving pretest contributed meaningfully to the variance in the first stage (at $p < .25$), then pretest was treated as a covariate in the second stage. If the effect of the pretest covariate was not significant, pretest was dropped from the analysis.
3. In analyses involving problem order, if no main effects or interactions involving that variable contributed meaningfully to the variance in the first stage of analysis (at $p < .25$), data from the two problems were pooled in the second stage.

This strategy enabled us to investigate two questions with respect to prior skill. First, did the effect of training differ as a function of prior skill? This was the case where the pretest score and treatment interacted in the first stage of analysis. Second, was preexisting skill a reliable predictor of performance independent of training? This was the case when pretest was a significant covariate in the second stage of analysis.

Hypothesis 1: The number and quality of arguments generated.

We predicted that trained participants would generate more and/or better arguments than controls when asked to explain the strengths and weaknesses of the given assessment (in the first question on each problem part). We turn first to a test of the number of arguments participants generated.

The number of arguments generated. To evaluate the effects of training on the number of arguments participants generated, we first parsed the officers' evaluations into constituent statements. Every statement was counted as an argument unless it summarized a judgment (e.g.,

"Assessment not good."), proclaimed ignorance (e.g., "I don't know"), was uninterpretable, or was illegible. This parsing procedure was executed on all responses to the six, two-part problem statements most frequently completed by experimental participants².

For example, problem 14A described the relationship of Arisle's people to their government and gave, as an assessment, that "The great majority of the population does not support a Mainlandia take over. They can be expected to generally support an invasion force." The following response to question one on problem 14A was parsed into four statements, as indicated by the numbers added by the authors:

(1) Assessment not good. (2) No reason to believe that governor would change after takeover by Mainlandia. (3) Furthermore, although they may back the governor they may not support the invasion force. (4) It also does not take into account the ethnic background and history of the island.

Three of these statements were counted as arguments (items 2-4). One statement (item 1) summarized a judgment, and thus was not counted as an argument.

For each participant, we computed the average number of arguments per posttest problem part. The posttest score served as the dependent variable in the full regression model. Independent variables were the treatment (training vs. control), problem order (order X vs. order Y), the average number of pretest arguments per problem part, and all interactions³.

In the full model there were no interactions involving pretest or problem order at $p < .25$. Thus pretest interactions were dropped and pretest score was retained as a covariate. Problem order was also dropped from the model, thus pooling data over the two. Thus, pretest interactions were dropped and only the main effect of pretest was retained, as a counterbalanced orders.

While trained participants generated more arguments on the posttest than did controls, the difference was not significant ($F_{1,33} = 2.377, p = 0.133$). The pretest scores reliably reduced error variance in the ANCOVA model ($F_{1,33} = 34.182, p < .0001$)⁴. Specifically, trained officers generated an average of 2.307 arguments per problem part on the posttest (2.248 after adjusting for pretest scores), while controls generated 1.567 arguments (1.733 after adjustment) (see Figure 2).

Training might have increased the number of arguments generated on one problem part more than the other. We tested for such effects in a more refined model in which problem part (the average number of arguments generated over part A and part B of all posttest problems) was a within-subjects variable. As before, between-subjects independent variables were

²All analyses concerning arguments were, thus, performed using data from six problems, each consisting of two problem parts. These were the analyses on argument quantity and quality. All other analyses were performed using data from all eight problems (each in two parts) executed by control and trained officers.

³Posttest arguments for one trained officer were inadvertently not examined by the SME, thus, that subject was dropped from analyses of the number and quality of arguments.

⁴The Pearson correlation between pretest and posttest scores was $r = .718, p < .001$.

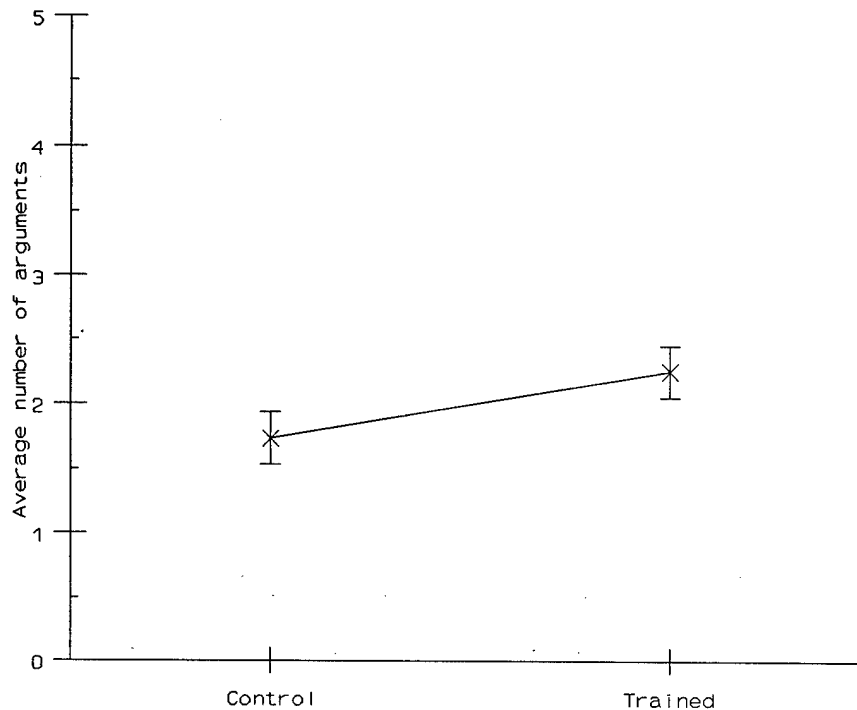


Figure 2. There was no statistical difference between trained officers and the control group on the number of arguments generated.

treatment, problem order, pretest, and all interactions.

Results of the full model included interactions involving pretest that were significant at $p < .25$, but no significant main effects or interactions involving problem order. Thus, pretest interactions were retained and problem order was dropped from the model.

Again there was no significant main effect for the R/M training to increase the number of arguments generated on the posttest ($F_{1,32} = 2.708, p = 0.110$). But, there was a significant interaction of treatment with pretest ($F_{1,32} = 5.941, p = 0.021$). Training conferred the greatest benefit on officers who had produced the most arguments on the pretest (see Figure 3). Only the coefficient for the regression curve representing trained officers was significant in simple regression, ($t_{26} = 6.265, p < .001$).

The more refined model exhibited no interaction involving treatment and problem part, despite the slight appearance that more arguments were generated on A problem parts than on part B (see Figure 4).

The quality of arguments generated. In order to evaluate the quality of participants' arguments, we solicited quality ratings about those arguments from the two subject matter

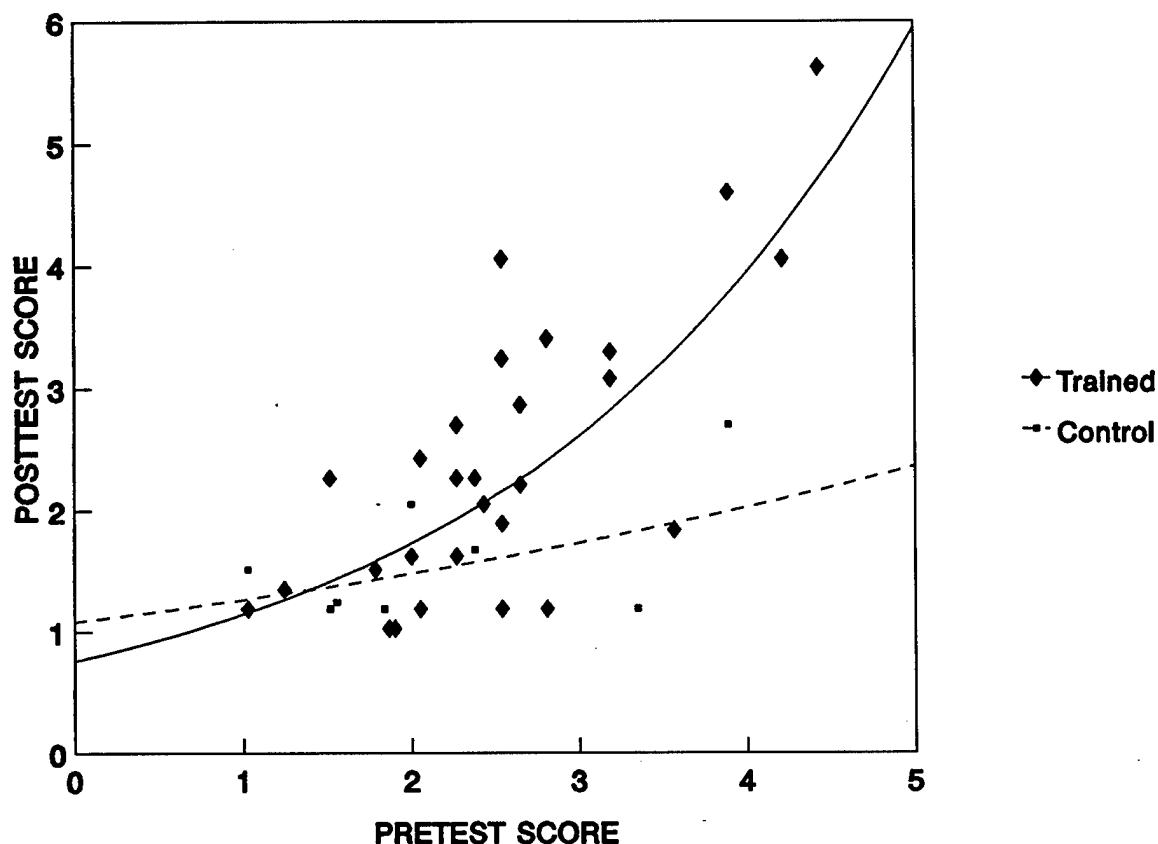


Figure 3. The largest number of posttest arguments was generated by trained officers who produced the most pretest arguments. The lines represent power regression curves for each group of participants.

experts (SMEs). The SMEs were given the test materials and arguments, parsed by the experimenter as described above. We asked them to rate each argument on a 5-point scale where 1 = very weak, 2 = weak, 3 = neutral, 4 = strong, 5 = very strong. The SMEs independently rated a small sample of arguments, discussed their ratings, and then rated the remaining items. SMEs were blind to the identity of each argument's author, the treatment the participant received, the remainder of the response of which the argument was a part, and whether the response was elicited on the pretest or posttest. The SMEs rated a total of 853 arguments. One SME rated 458 arguments, one rated 484, and 89 arguments were rated twice, either by two different raters, or by the same rater. Interrater reliability for the 40 items rated by two different judges was moderate (Pearson's $r = .534$, $p < .001$). Prior to analysis, quality ratings for arguments judged by both SME's were averaged.

For each participant, we computed the average quality score per argument on each test. The posttest score served as the dependent variable in a model in which the independent variables were treatment, problem order, the average pretest quality score, and all interactions. This model exhibited no interactions involving pretest at $p < .25$, nor any effects involving problem order at $p < .25$. Thus, pretest interactions were dropped and data from the two problem orders were pooled. Pretest exhibited no reliable main effect in the resulting ANCOVA, and so it, too, was dropped. The resulting model was a simple ANOVA, in which treatment was the sole factor.

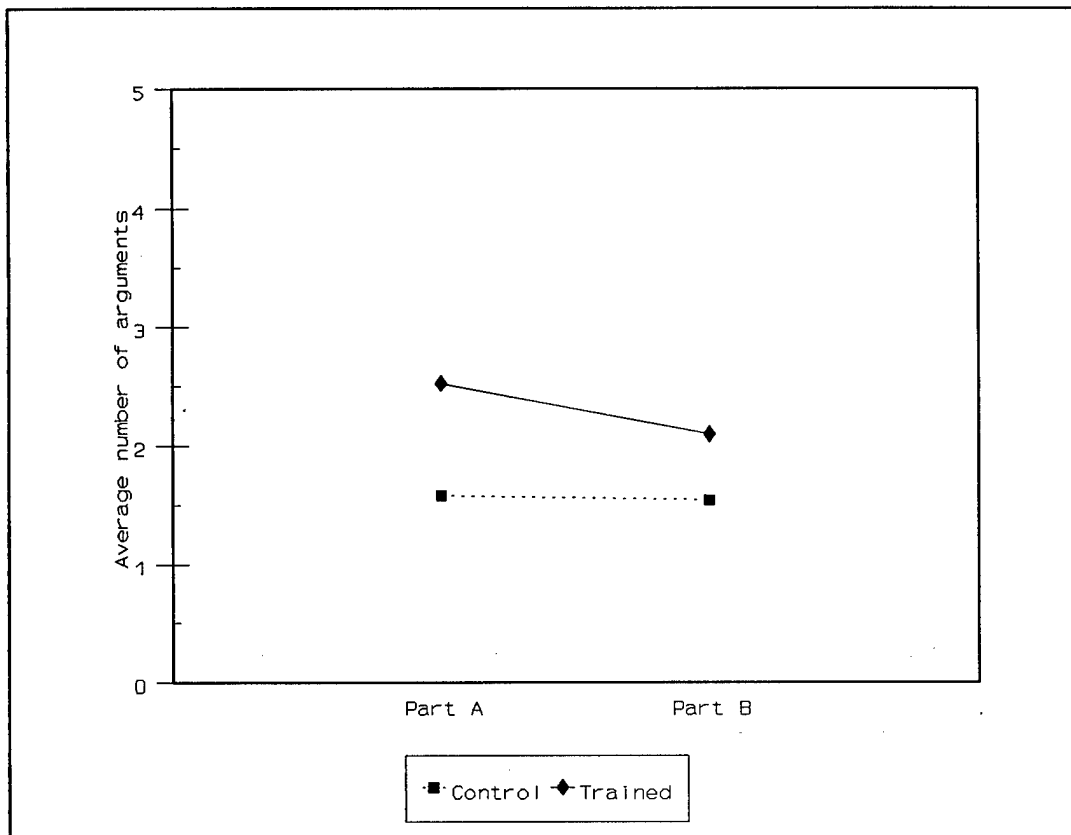


Figure 4. Trained officers appeared to generate more arguments than controls on both problem parts but the difference was not significant.

Training did not raise argument quality significantly ($F_{p,34} = 3.759, p = 0.061$). The mean argument quality score for controls was 2.751, and for trained officers argument quality was 14% higher: 3.128 (see Figure 5). Also there were no significant interactions of training with problem part on argument quality.

The sole finding of interest was an interaction between treatment and the pretest control score. Training tended to increase the number of arguments generated on the posttest for the trained officers relative to the control group for those who made more arguments on the pretest.

Hypothesis 2: Generating conflicting arguments.

There was no difference between trained participants in the quality of arguments they produced over controls. We looked deeper to see if quality would have a stronger effect by considering the relevancy of the arguments, i.e., arguments that either support or disconfirm the assessment, as compared to irrelevant or neutral arguments. In particular, we had predicted that training would enable participants to notice more evidence that conflicted with an assessment, and that they would capitalize on this ability by generating more arguments that disconfirmed the given assessment.

To test this prediction, an experimenter coded each argument that participants generated

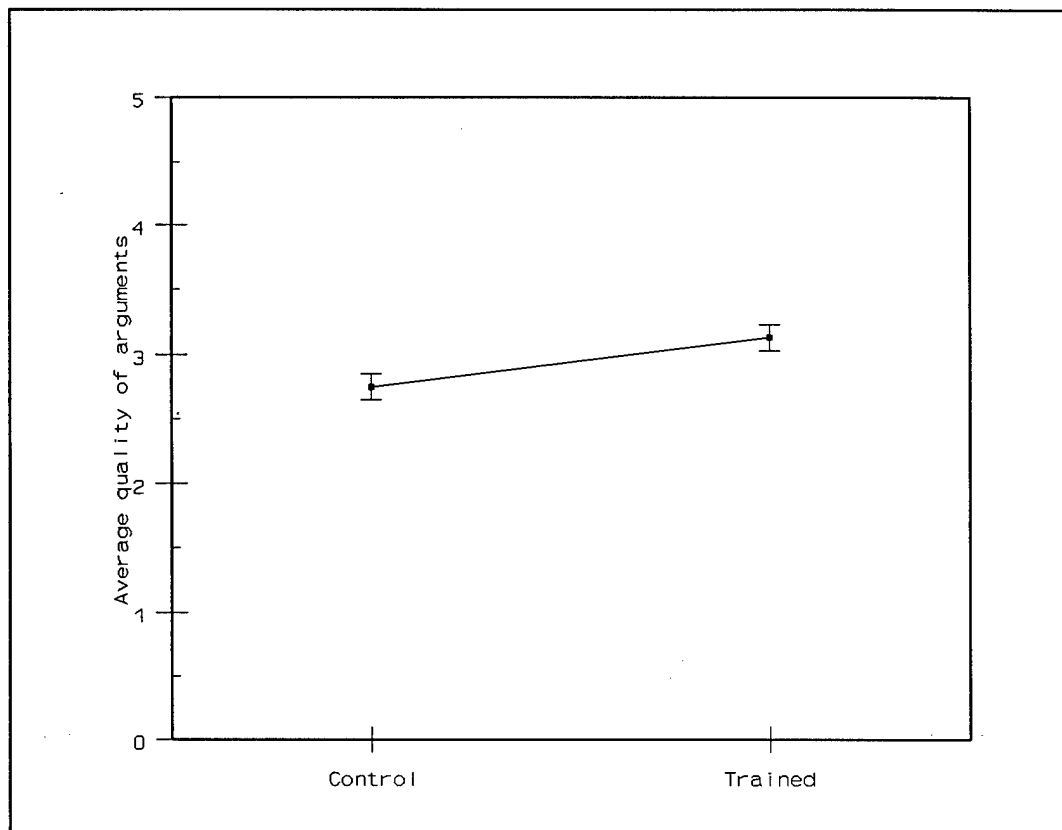


Figure 5. There was no significant difference between trained officers and the control group on the overall quality of arguments generated during the posttest.

as either disconfirming the assessment, supporting it, or standing neutral with respect to it. Subject matter experts reviewed and corrected this coding of argument direction.

As illustrated in Figure 6, trained officers generated more disconfirming statements on the posttest than did controls (44% of 334 statements vs. 37% of 74 statements by controls), slightly more supporting statements (31% vs. 24%), and markedly fewer neutral statements (25% vs. 39%). The interaction of treatment and argument direction was significant in a chi-square test that crossed treatment conditions (2) with argument directions (3) ($\chi^2_2 = 6.577, p = .037$). Trained officers generated more relevant arguments (i.e., supporting and disconfirming arguments combined) on the posttest than neutral arguments, relative to controls ($\chi^2_1 = 6.555, p = .010$). However, training did not predispose officers to generate a significantly greater proportion of disconfirming arguments, relative to other arguments. Nor did training help officers produce a significantly larger proportion of supporting arguments, relative to other arguments.

We next examined the potential influence of problem part on the distribution of arguments by direction. On both problem parts, a smaller proportion of the arguments of trained officers were neutral than was the case for controls. However, the trained officers distinguished themselves on part A by generating a larger proportion of *disconfirming* arguments (45%) than did controls (27%), while on part B trained officers generated a larger proportion of *supporting* arguments (33%) than did controls (16%). (See Figure 7). To test the significance of

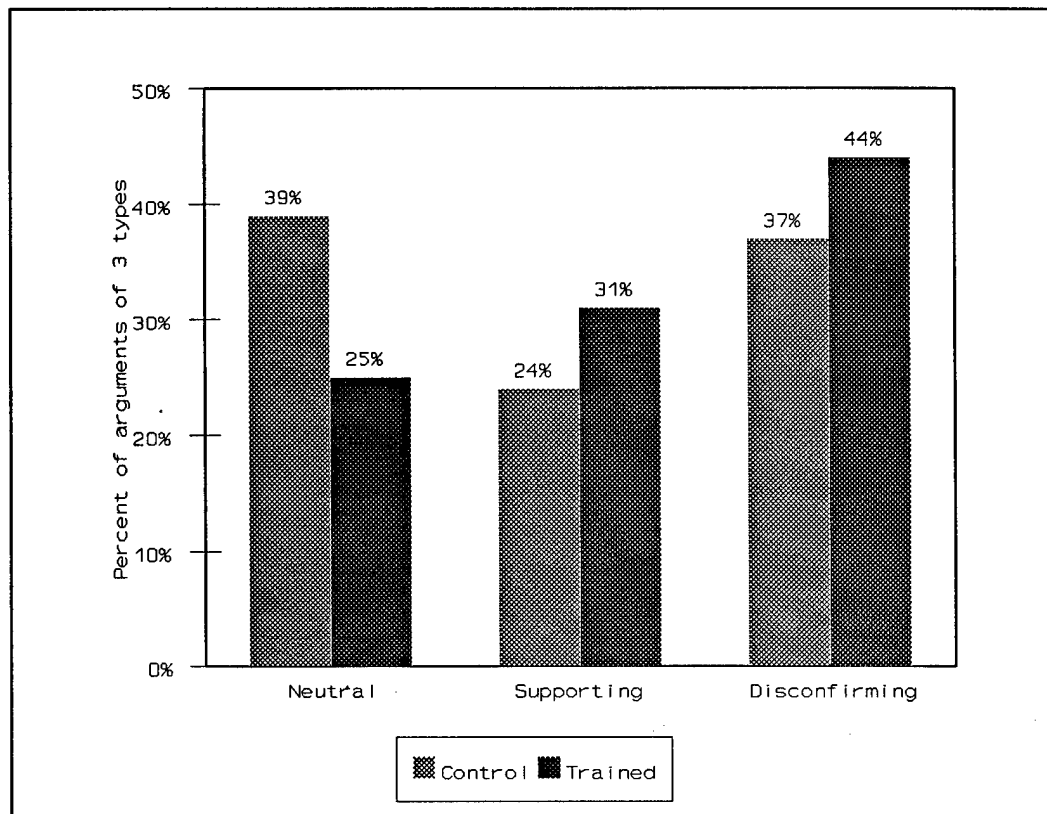


Figure 6. Trained officers generated significantly fewer neutral statements after training. Trained officers generated arguments that were more relevant, either supporting or disconfirming an assessment.

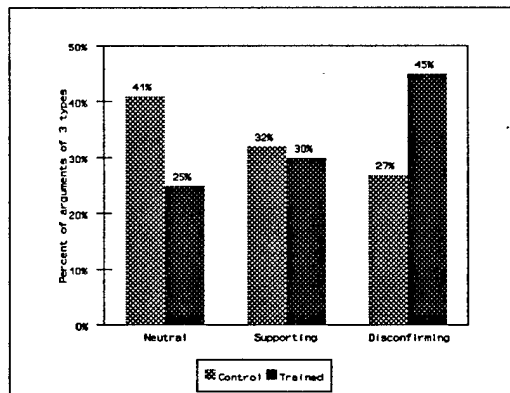


Figure 7a. The graph presents the proportion of arguments of each type generated by participants in both training conditions on part A.

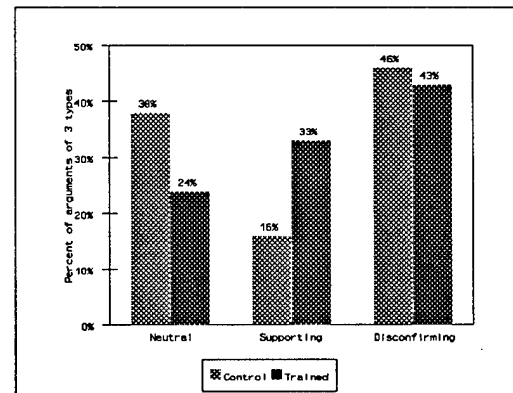


Figure 7b. The graph presents the proportion of arguments of each type generated by the participants in both training conditions on part B.

this apparent interaction of treatment and direction with problem part, a log-linear model was constructed consisting of all two way interactions between treatment, problem part, and direction and their main effects, but omitting the three-way interaction.⁵ The χ^2 statistic for this model

⁵Omitting the interaction of treatment, direction, and problem part tested the independence of those three variables.

was not significant ($\chi^2_2 = 4.09, p = .129$). Thus, there was not a significant difference in the distribution of arguments by direction between groups and problem parts.

Hypothesis 3: Effects of training on assessments and accuracy.

The bottom line in evaluating the effect of training on situation assessment is whether the instruction improved accuracy. Officers who are able to interpret battlefield events correctly (and in a timely manner) are presumably more likely to achieve victory. Those who do not are more likely to suffer defeat. We predicted that training would increase accuracy. In testing this prediction, we first ask whether there was any effect of training on assessments at all. We examined the impact of training on participants' own numeric evaluation of each assessment, which they provided in the form of ratings of agreement with the assessment at each problem part. Next, we examined whether such effects increased the accuracy of participants' agreement ratings, relative to the agreement ratings by the most senior of our SMEs.

Participants were asked to evaluate the assessments on a scale, where -2 indicated extreme disagreement, 0 neutrality, and 2 extreme agreement. Thus, scores greater than zero represent degrees of agreement with the assessment, while scores less than zero represent the corresponding degrees of disagreement with the assessment.

A regression model was constructed for each problem, in which the dependent variable was the average agreement rating per problem part for the given problem on the posttest. Independent variables were treatment and average agreement rating over all pretest problem parts.

Effects on assessments. Ratings by trained officers were closer to those of the SME on problems 5, 9, 10, 12, and 14. On two problems (1, 2) mean agreement ratings were virtually identical between groups. On one problem (3), mean agreement ratings by controls were slightly closer to the SME than were the ratings by trained officers.

On problem 5, we removed pretest interactions from the model and, finding that pretest did not account for significant variance in ANCOVA, removed it from the model altogether. In the resulting ANOVA, there was a significant main effect of treatment ($F_{1,16} = 5.487, p = 0.032$). The mean agreement rating of controls (0.000) differed significantly from that of trained officers (-1.036), which in turn lay closer to the mean SME rating of -1.5 (-1 on part A, -2 on part B).

For all other problems, the full model reduced to ANOVA, and main effects of training were not significant.

Did treatment effects vary by problem part? Inspection of the mean agreement ratings indicated that the groups performed similarly with respect to one another on parts A and B. On all but one problem, the ordinal relationship between groups was the same on both problem parts. For example, where trained officers rated an assessment higher than controls on part A, they also rated it more highly on part B.⁶

⁶The exception was problem 1, on which trained officers produced a higher rating (-.370) than controls (-1.000) on part A, but virtually the same rating (-.583) as controls (-.500) on part B.

To investigate the interaction of problem part with training, we replaced the single dependent variable in the full model with a within-subjects variable, the agreement rating on part A and part B, respectively. Problems 5 and 12 produced interactions involving test and problem part. In both cases, interactions involving pretest (at $p < .25$) led us to use the full regression model.

On problem 5, trained officers were more accurate than controls on both problem parts, with respect to the SME agreement ratings of 2 on part A and 1 on part B. (See Figure 8). In addition, there was a significant interaction of treatment with problem part ($F_{p14} = 5.801$, $p = 0.030$). Training had more impact on agreement (and accuracy) in part B than in part A.

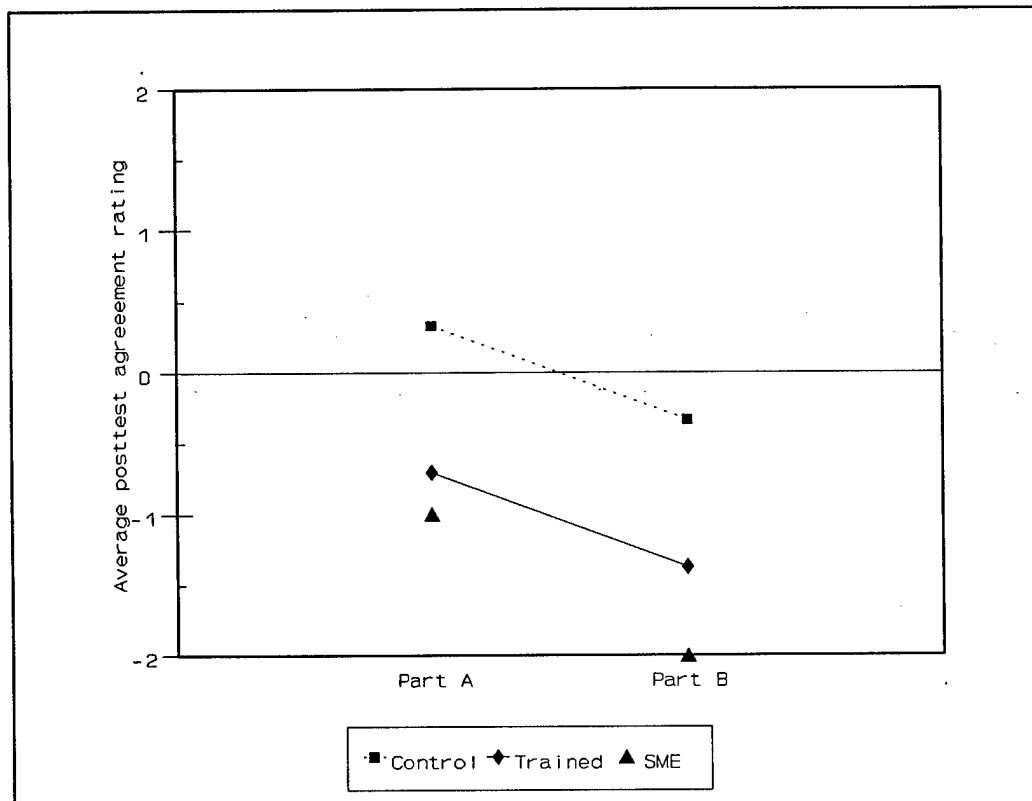


Figure 8. Mean agreement ratings were lower for trained officers than controls on problem 5, lower on part B than on part A, and differed more between groups on part B.

On problem 12, similarly, mean ratings by trained officers were closer than controls' ratings to those of the SME on both problem parts. (The SME gave both parts an agreement rating of 2). (See Figure 9). There was a significant treatment by problem part interaction ($F_{p13} = 7.199$, $p = 0.019$). Once again, training had more impact on agreement (and accuracy) in part B than in part A.

Effects on accuracy. Training had an effect on assessments in two of the problems. In addition, these effects appeared to be consistently in the direction of greater agreement with the SME. We can test the latter relationship more directly. A convenient measure of error is the absolute value of the difference between the SME's rating and the participant's rating on a given

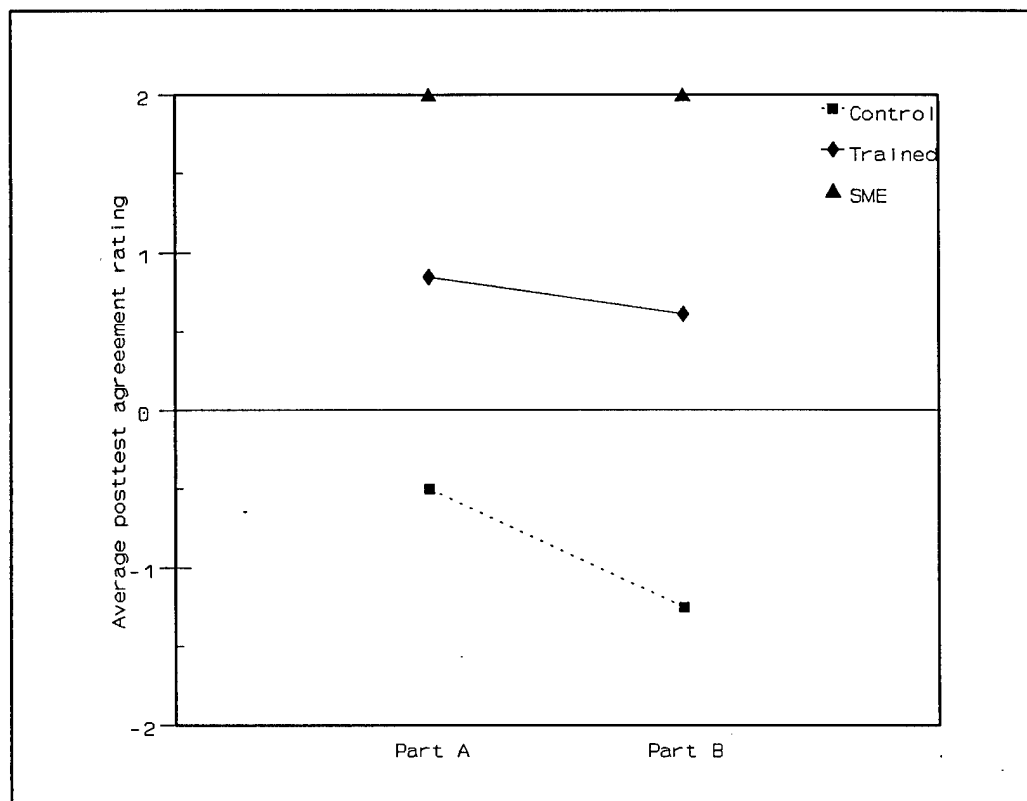


Figure 9. Ratings by trained officers were higher than those by controls, and thus closer to the SME's rating of 5, on both parts of problem 12.

problem part. For convenience in exposition, we can convert this measure of error to a measure of accuracy by subtracting it from the maximal value on the agreement scale (5). Thus, an error of zero corresponds to an accuracy of 5.

Average accuracy scores were higher among trained officers than controls on problems 5, 12, and 14, higher among controls than trained officers on problem 3, and similar between groups on the remaining problems.

We constructed a model for each problem, in which the average accuracy rating over problem parts served as the dependent variable, and independent variables were treatment, group, the average accuracy over all pretest problem parts, and all interactions of these variables.

Trained officers were significantly more accurate than controls on problems 14 and 12. On problem 14, the mean accuracy score of trained officers was 19% higher than that of controls (4.321 vs. 3.625; $F_{1,16} = 6.049$, $p = 0.026$). On problem 12, the score of trained officers was 76% higher than that of controls (3.731 vs. 2.125; $F_{1,15} = 5.145$, $p = 0.039$). (ANOVA output was used for these problems because pretest interactions failed to achieve $p < .25$ and pretest did not account for significant variance in ANCOVA.) Accuracy on the pretest did not predict accuracy on posttest problems. Pearson correlations between pretest and posttest scores were $r = .108$, $p = .679$ for problem 12; $r = -.167$, $p = .508$ for problem 14.

To explore potential interactions of the treatment with problem part, the model was refined to include problem part as a within-subjects variable, comprising the accuracy scores on part A and part B of the given problem. The between-subjects variables remained the same as in the previous model: treatment, accuracy over all pretest problem parts, and their interactions.

Significant treatment by problem part effects were found on problem 5. Trained officers had higher accuracy scores than controls on part A (4.429 vs. 3.75) and part B (4.357 vs. 3.337) (see Figure 10). There was a reliable interaction of treatment with problem part ($F_{1,14} = 5.657$, $p = 0.032$). Training produced a larger increase in accuracy in part B than part A.

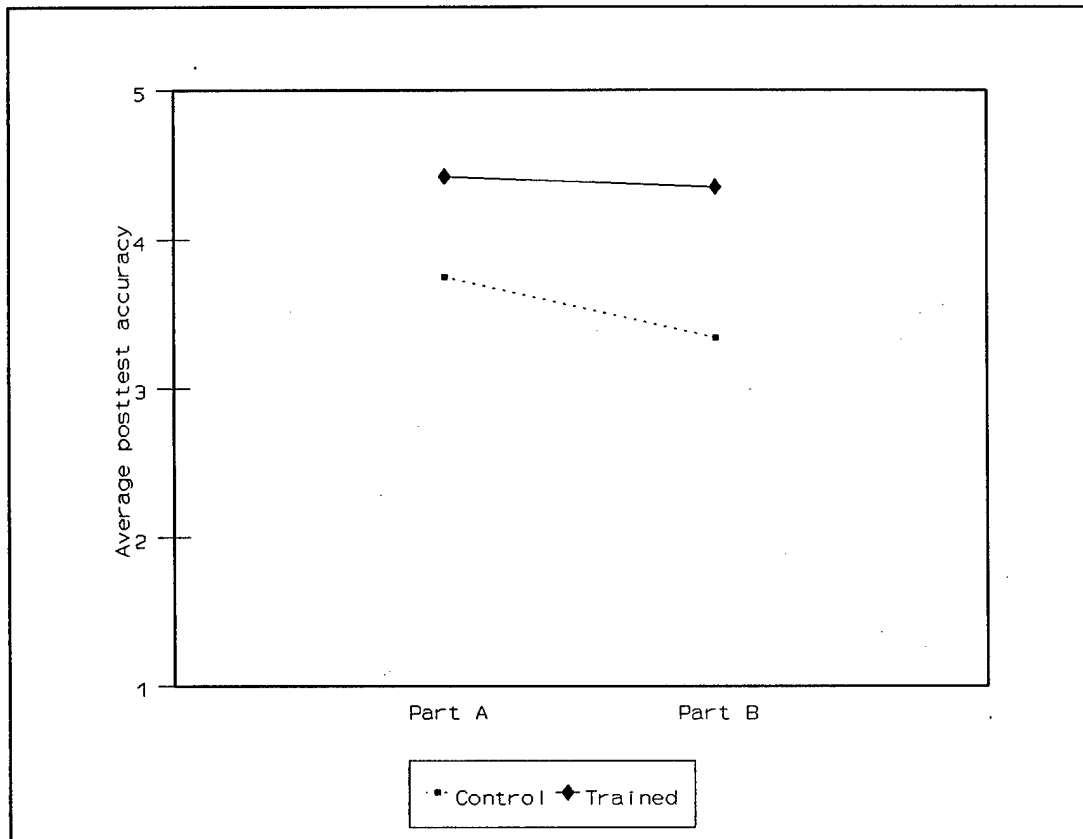


Figure 10. On problem 5, trained participants and controls differed more on problem part B than problem part A.

Discussion. Training caused the officers to generate more relevant arguments on some problems when they evaluated assessments. Training also increased the number of arguments generated. It is a somewhat surprising finding nonetheless, because the training was so brief, lasting only 90 minutes. It was entirely possible that ingrained habits of reasoning would overwhelm the short instruction and practice in these particular critical thinking skills. In the educational literature, the introduction of a new problem-solving method is often found to handicap more experienced participants, presumably because it conflicts with over-practiced techniques (Cronbach & Snow, 1977; Lajoie, 1986). This did not occur here. Officers were able to put the new techniques to work immediately. There was an indication (in the analysis of argument quantity) that officers with greater prior skill, i.e., higher pretest performance, benefitted most from the instruction.

The influence of training on argument quality was at least partly due to changes in the direction of arguments. Training increased the number of arguments officers made that either supported or disconfirmed the given assessment, while it reduced the production of neutral or irrelevant arguments. Training helped officers to formulate more persuasive arguments and fewer inconsequential ones. Much of the training focused on detecting conflicting evidence and, thus, it was quite possible that there would have been a higher proportion of disconfirming arguments. This did not occur. Officers were instead quite balanced in their application of the trained critiquing skills. They used them to generate relevant arguments for *and* against the given assessment, rather than to construct neutral arguments.

Finally, training improved the accuracy of participants' assessments on several problems. Raw agreement ratings differed between groups on problem 5. Training increased accuracy on this problem. The agreement ratings by trained officers were significantly closer than those by controls to the ratings of an SME. Analyses on accuracy were even more striking. When closeness to the SME rating was used as a measure of accuracy, trained officers were significantly more accurate than controls on problems 12 and 14. Hypothesis 3 was supported.

We wish to draw particular attention to the strong effect of treatment on accuracy in problem 14. Prior to conducting the analysis of agreement ratings, we had asked the senior SME to predict potential effects of training on responses to several test problems. The SME reported that one problem, number 14, was dramatically more likely than any other to illustrate the effects of training on accuracy. As he stated, "This one [problem] lends itself more than any other to the positive impact of training - questioning assumptions underlying the assessment." The results on this particular problem support his prediction.

In sum, brief training in selected metacognitive skills improved the quality of arguments officers generated when evaluating situation assessments, and it did not lower the number of arguments that officers conceived. Simply put, training made officers more efficient at evaluating assessments. The improvement in quality was due in part to changes in the type of arguments officers generated. Trained officers produced more disconfirming and supporting arguments than neutral arguments, relative to controls. That is, training enabled officers to generate more arguments that could make a difference. Finally, training improved the accuracy with which officers evaluated the given assessments. That effect was particularly significant on the one problem that an SME predicted would elicit strong differences between groups.

Exploration of Research Questions

Question 1. What is the effect of training on the so-called confirmation bias?

Recall that participants rated their agreement with an assessment in part A of each problem. In part B they read new information containing some elements that supported the given assessment, some elements that disconfirmed it, and some neutral assertions, and rated their agreement with the assessment again. Theories of confirmation bias would predict that in part B participants would give more weight to elements of new evidence that supported their prior view of the assessment. The confirmation bias, however, is a special case of a more general possibility: the influence of prior opinions on the impact of new information. The opposite effect is also possible, in which participants give more weight to new information that disconfirms their original hypothesis (the "disconfirmation bias"). We were interested in whether either of these patterns appeared in the data, and the influence of training on these effects if they were present.

We constructed a model in which the dependent variable was agreement ratings for the A and B parts of a given problem. Problem part was a within-subjects variable. There were three between-subjects independent variables. Agreement was a dichotomous variable indexing whether the participant agreed or disagreed with the assessment prior on part A. If a participant gave a neutral rating of the assessment on part A, his or her response was discarded from the data set for analysis of that problem. The remaining independent variables were treatment and an index of pretest agreement, computed by averaging the rescaled agreement ratings over all pretest problems and problem parts.

The model was run separately on data from each of the eight problems executed by both groups. For three problems (1, 3, and 14) no model could be run because there were no participants in one of the four cells of the between-subjects design (treatment (2) x agreement (2)). Pretest did not have a significant effect in any of the remaining problems, and several of the full models generated singular matrices. Thus, pretest was dropped from all models, and ANOVAs were used.

The first issue of interest is a possible main effect of problem part. If agreement ratings consistently rise or fall after examination of new evidence on a given problem, participants agree on the interpretation of the new evidence as confirming or disconfirming the stated assessment. Data from four of the five problems (2, 5, 9, and 12) exhibited a main effect of problem part in which agreement ratings declined between problem parts over all participants. This indicated that participants interpreted the new information in part B as disconfirming the given. There was no effect of problem part in problem 10. There were no significant interactions.

Biases in responding to new data are reflected in interactions between problem part and agreement. For example, if officers who agree with the stated assessment in part A agree with it more in part B, while those who disagree with the stated assessment in part A disagree with it more in part B, there is evidence for a confirmation bias. Regardless of what their part A assessment is, each group interprets the new evidence as supporting it. (In the graphs, this shows up as two lines diverging away from the neutral line.) On the other hand, if officers who agree with the stated assessment in part A agree with it less (or even disagree with it) in part B, and officers who disagree with the stated assessment in part A disagree with it less (or even agree with it) in part B, there is evidence for a disconfirmation bias. Regardless of their part A assessment, each group interprets new evidence as disconfirming it. (In the graphs, this shows up as two lines converging toward the neutral line or crossing over the neutral line.)

Data from three problems exhibited an interaction of problem part with agreement that signified a "disconfirmation bias," i.e., a tendency to give more weight to new evidence that conflicts with one's earlier assessment. Participants who agreed with the assessment on part A interpreted new information on part B as disconfirming the assessment, while those who initially disagreed with the assessment did not change their agreement ratings in response to the new information. This effect appeared on problems 5 ($F_{1,12} = 8.526$, $p = 0.013$), 9 ($F_{1,13} = 6.927$, $p = 0.021$), and 12 ($F_{1,13} = 5.157$, $p = 0.044$) (see Figure 11).

The pattern of change for problems 9 and 12 would be an indication of training decreasing the disconfirmation bias (see Figure 12), although these effects were not significant. Problem 10 revealed a quite different pattern (see Figure 13). In this problem controls tended toward a confirmation bias and trained subjects toward a disconfirmation bias. The interaction of treatment with problem part and agreement was significant ($F_{1,12} = 5.479$, $p = 0.037$).

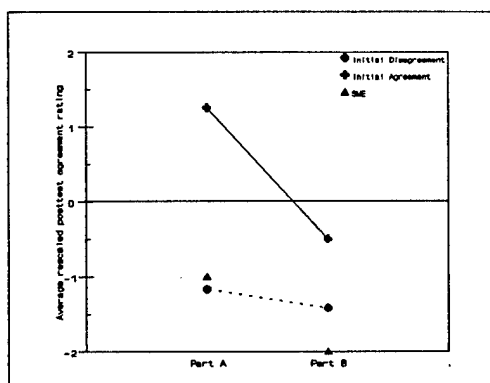


Figure 11a. Officers who initially agreed with the assessment for problem 5, interpreted new information as disconfirming (solid line).

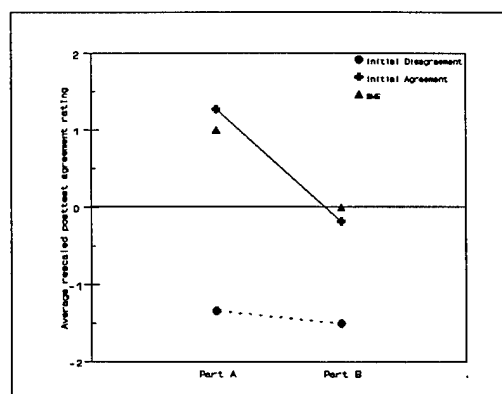


Figure 11b. Officers initially agreeing with the assessment on problem 9, interpreted the new information as disconfirming (solid line).

These results seem contradictory. In problems 9 and 12, the pattern seems to show that training appeared to *reduce* the disconfirmation bias, while in problem 10 it appeared to reduce the confirmation bias and to *increase* the disconfirmation bias. Although only results for problem 10 was significant, we are inclined to take the inconsistency with problems 9 and 12 seriously. The effects of training should not be understood in terms either of reducing the confirmation bias (problem 10) or reducing the disconfirmation bias (problems 9 and 12). In all three of these problems, changes induced by training caused officers to be more accurate than controls in part B (relative to the assessments of the SME). Thus, trained officers were less likely to explain away conflicting data when the explanations would have been implausible (problem 10). But they were more likely to explain away conflicting data when the explanations would have been plausible (problems 9 and 12). The unifying pattern underlying these results is the more accurate *evaluation of explanations* as a result of training.

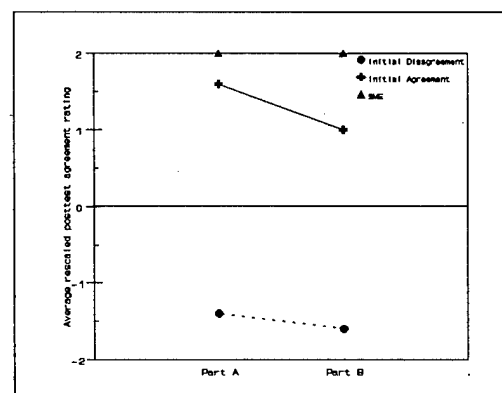


Figure 11c. Officers initially agreeing with the assessment on problem 12, interpreted new information as disconfirming the assessment (solid line), while those originally disagreeing did not change.

Question 2. How does training affect the sensitivity of participants to new evidence?

Providing officers with new tools for evaluating information had the potential to make them more or less sensitive to aspects of a given problem, relative to controls. For example, trained officers might change their initial agreement ratings (e.g., Part A in this research) more than controls in response to new information on part B of each problem, or they might shift their ratings less. Training that produced large shifts in sensitivity would be suspect, given that the participants in this study were experienced officers whose judgments were presumably fairly well-calibrated. Small but significant differences in sensitivity between the treatment groups might be interesting pointers to future research.

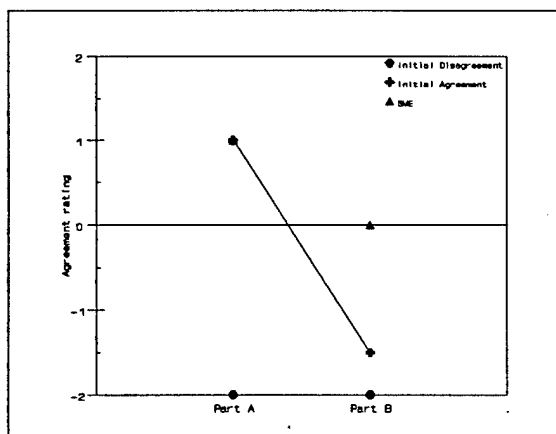


Figure 12a. The controls initially agreeing display a disconfirmation bias on problems 9.

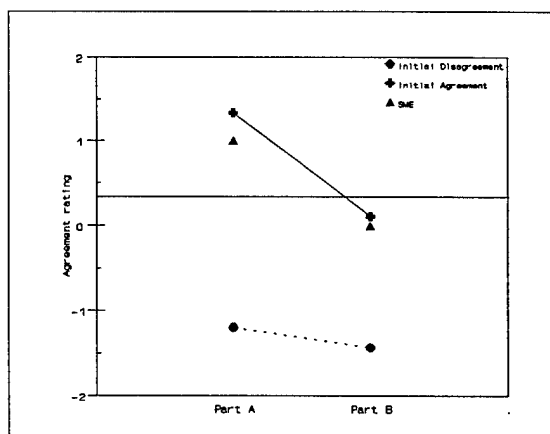


Figure 12b. Training appeared to moderate the disconfirmation bias on problem 9.

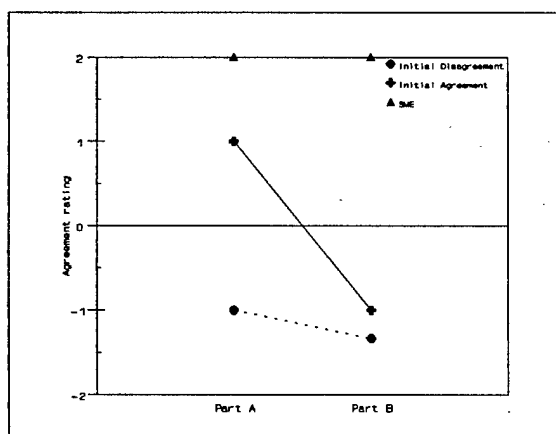


Figure 12c. Control officers initially agreeing with the assessment on problem 12 exhibited a disconfirmation bias.

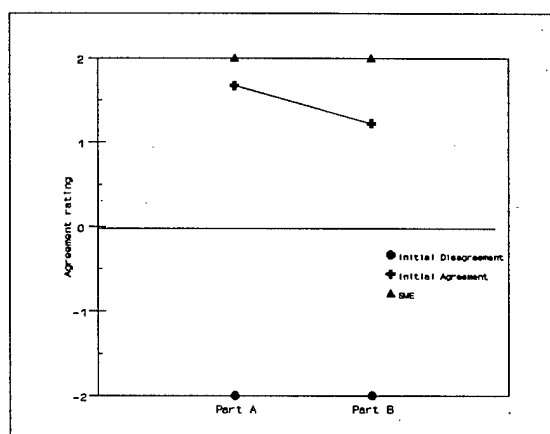


Figure 12d. Training appeared to moderate the disconfirmation bias on problem 12.

The metric of sensitivity in these analyses was the absolute value of the agreement rating on part B (after receiving new information) less the rating on part A. Higher scores indicated greater sensitivity to the new information on part B. For each participant, a sensitivity score was computed for each posttest problem, and a pretest sensitivity measure was computed by averaging sensitivity over all pretest problems.

Mean sensitivity scores exhibited no clear pattern over all problems. The scores varied between zero and one for trained officers on every problem, and for controls on all problems except 5 and 9. The mean sensitivity of controls was higher than that of trained officers on four problems (2, 5, 9 and 12), lower on two (3 and 10) and the same on the remaining two problems (1 and 14).

Overall effects of treatment on sensitivity were tested in a model that had as its dependent measure the average sensitivity score over all problems. Independent variables were treatment, problem part, pretest, and their interactions. No significant effects of treatment were found in analysis of variance.

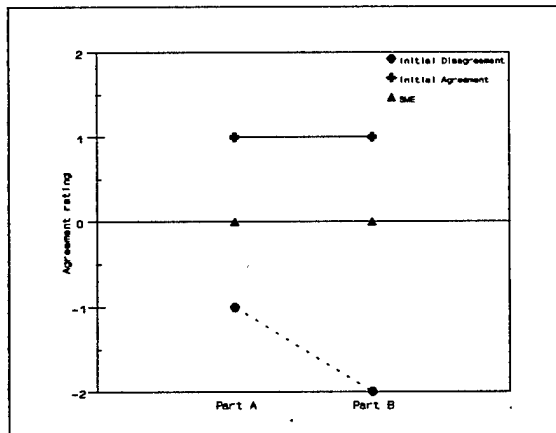


Figure 13a. Controls tended toward a confirmation bias on problem 2.

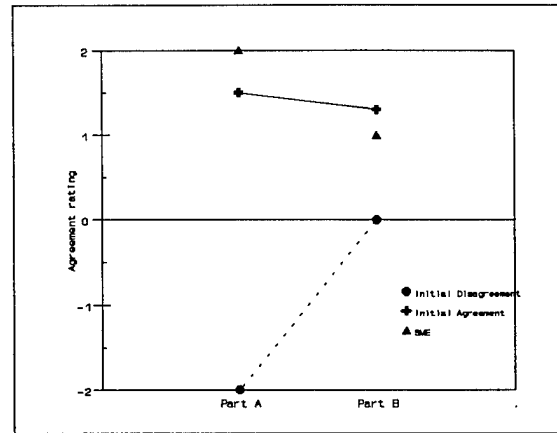


Figure 13d. Training appeared to reduce confirmation bias on problem 10.

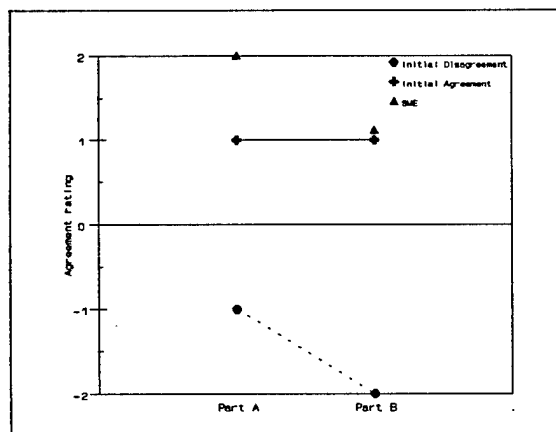


Figure 13c. Controls tended to a confirmation bias on problem 10.

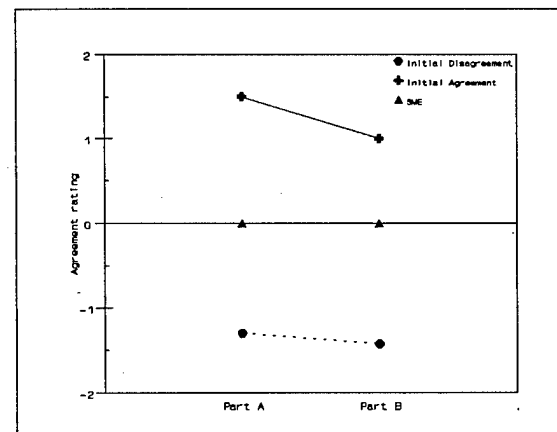


Figure 13b. Training appeared to reduce confirmation bias on problem 2.

Analyses of each problem were undertaken using a model in which the posttest sensitivity score served as the dependent variable, and the independent variables were treatment, pretest, and their interaction. Only data from problem 12 exhibited significant effects involving treatment ($t_{13} = 2.278, p = .040$) when the full model was interpreted. On that problem, mean sensitivity to new information was .75 among controls, and .385 among trained officers. We noted in the previous section that the response of controls to new information on problem 12 made their assessments less accurate. Training had little to no effect on over-sensitizing participants to new information.

Question 3. What is the effect of training on confidence in assessments?

The training was explicitly intended to alter the way officers evaluated situations, and the change in cognitive strategy might in turn have affected their confidence in their decisions. It seemed possible that trained officers would have less confidence in the decisions they made using the newly learned techniques. Lower confidence might be expected given the emphasis in training on exploring alternative assessments and identifying conflicting evidence. Lower confidence might also be induced by the need to utilize unfamiliar decision making techniques.

On the other hand, training might also improve confidence. It provides tools for handling ambiguous and conflicting evidence, and emphasizes the need to make timely decisions despite such uncertainty.

Confidence is reflected in the distance of an agreement rating from the neutral midpoint of the agreement scale. Thus, as a metric of confidence, we took the absolute value of each rescaled agreement rating (i.e., agreement less the scale's midpoint of 3). If training changed officer's confidence, we would expect that the confidence score would be reliably larger or smaller among trained officers than controls.

The mean posttest confidence scores on each problem (averaged over both problem parts) were characterized by high variance within groups and similar means between groups and between problems. Means ranged from .5 to 1.5. Only on problem 14 were group means clearly different, and on that problem, controls were more confident than trained officers.

Further analysis by problem employed a model in which the dependent variable was the average confidence score over the two parts of a given problem. The independent variables were treatment, pretest, and their interaction. ANOVA output was interpreted for all problems because pretest did not reliably account for error variance either in interaction with other variables in the full model, or as a main effect in ANCOVA.

Training produced a significant effect only on problem 14, on which the mean confidence of trained officers (.679) was 51 percent lower than that of controls (1.375) ($F_{1,16} = 6.049, p = 0.026$)⁷. To explore possible interactions of the treatment with problem part, the full model was expanded. Two dependent variables, each a confidence score on one part of the given problem, represented a single within-subjects variable called problem part. The remaining independent variables were the same as in the previous model: treatment, average confidence across pretest problem parts, and interactions.

Only problem 5 exhibited an interaction of treatment with problem part. On that problem trained officers were less confident than controls on part A and more confident than controls on part B ($F_{1,14} = 7.446, p = 0.016$) (see Figure 14). The confidence of controls was possibly unfounded, given that trained officers matched the SME agreement rating better than the control group.

Question 4. Which sources of individual difference predict ability to evaluate assessments?

The officers who participated in this study varied on several dimensions of experience that potentially influenced their performance on the tests. These individual differences were:

Years of service -- Participants varied on their tenure in the military. The mean length of military tenure was 10.7 (SD = 5.03).

Rank -- Participants varied in rank. The officers included two 1st Lieutenants, 18 Captains, 16 Majors, and one Lt. Colonel.

⁷The Pearson correlation between pretest and posttest scores was $r = .455, p = .058$.

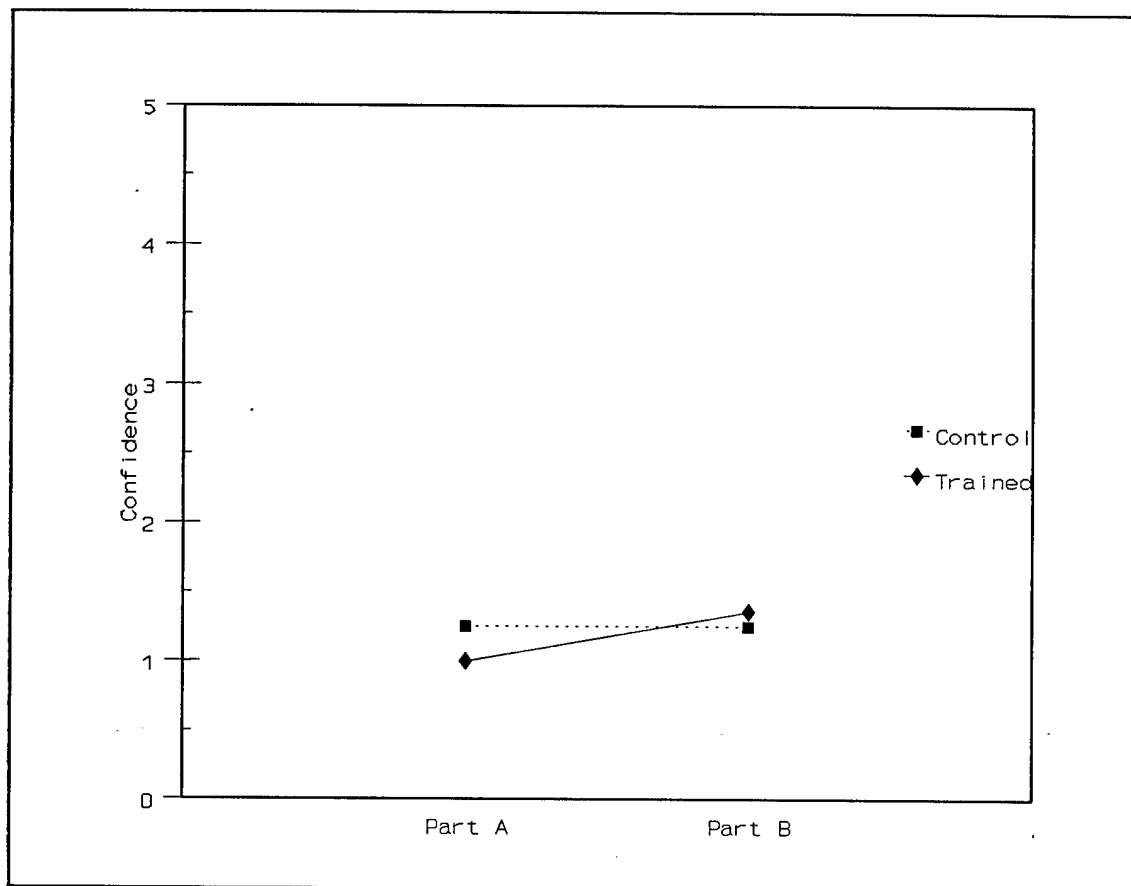


Figure 14. Confidence was lower among trained officers than controls on part A of problem 5 but significantly higher on part B.

Officer's training -- Participants varied on whether or not they had attended Command and General Staff College (CGSC). Fourteen had attended and 23 had not.

Positions -- The participants had performed their duties in a range of positions. Those we believed were most relevant to performance on the tests were assignments as S-3, S-3 staff, G-3, G-3 staff, executive officer (XO) at the battalion level or higher, and commander at the battalion level or higher. Twenty-six officers had accumulated an average of 27 months in such positions ($SD = 20$). Eleven officers had no experience in any of these positions.

We wished to learn whether any of these attributes reliably predicted the number or quality of arguments officers generated, agreement ratings, or scores on accuracy, sensitivity, or confidence. It is important to note that even strong effects of one or more of these attributes would not support inferences concerning causality. For example, an effect of position assignment on the quality of arguments might mean either that experience in a particular position enhanced abilities in situation assessment, or that stronger abilities in situation assessment lead to assignment in specific positions. Still, reliable effects of individual difference variables might raise issues for future research.

We coded the attributes in several ways. Years of service was coded as a continuous

variable (years) and as a dichotomous variable (less than the median ten years service vs. ten years of service or more). Rank was coded ordinally (1 through 4), from 1st Lieutenant through Lieutenant Colonel. Officer's training was coded dichotomously (took CGSC training vs. did not take the training). Experience in positions was coded dichotomously (experience in any one or more of the specified positions vs. no experience in any of the specified positions), continuously as the number of specified positions held (e.g., S3 = 1; S3 and Assistant G-3 = 2), and continuously as the total number of months in the specified positions.

Pearson correlations were then computed in order to identify strong relationships between each of the individual difference measures above and the average posttest scores (over all problem parts) on argument quantity, quality, agreement, and accuracy. No significant correlations were found between these performance measures and CGSC attendance or rank.

Significant correlations were indicated between accuracy and experience. For both independent variables, a full regression model was constructed whose independent variables were treatment, problem order, the individual difference measure, the relevant pretest score, and all interactions. We reduced the full model as necessary by eliminating unproductive terms, and in each case used an ANCOVA model consisting of main effects of treatment and the individual difference variable.

A reliable predictor of accuracy relative to SMEs was the number of months that officers had spent in positions typically associated with situation assessment activities (G3, G3 staff, S3, S3 staff, XO or commander). The correlation between the two measures was negative (*Pearson's* $r = -.439$, $p = 0.007$) indicating that high experience was related to better accuracy. In ANCOVA, the number of months in relevant positions was significant ($t_{34} = -2.668$, $p = 0.012$).

Discussion. The research questions addressed here concerned effects of training on decision bias, as well as the influence of individual difference variables on several performance measures.

In the analysis of decision bias, we found additional evidence for the improvement of accuracy by training. For example, new information in four of five problems was interpreted as disconfirming the given assessment. That is, participants on average lowered their agreement ratings after reading the information in part B.

In two problems an early assessment influenced the interpretation of new evidence. In both cases, the results fit the pattern of a disconfirmation bias, that is, placing more weight on new information that disconfirmed the initial assessment than on information that might have confirmed it. Training appeared to counter this bias, resulting in higher accuracy (but in neither case was the effect significant). In two problems, training had an opposite effect, reducing a confirmation bias (the tendency to place more weight on new information that confirms the initial assessment) and increasing a disconfirmation bias - in both cases, once again, increasing accuracy as a result. (One of these effects was significant.) Thus, there appears to have been little if any systematic influence of training on disconfirmation or confirmation biases.

A better view of these results is that trained participants learned effective and flexible strategies for handling evidence. Training taught them to consider explaining away conflicting evidence; hence, they were less susceptible to disconfirmation biases in problems 9 and 12 (see Figure 12). Training also taught them to generate alternative assessments; hence, they were less subject to confirmation biases in problems 2 and 10 (see Figure 18). Most importantly, training

taught them to step back and *evaluate* the explanations of conflicting evidence, and to step back and *evaluate* the alternative assessments. As a result, they explain away conflicting evidence when it is appropriate and plausible to do so (officers who agreed with the given assessment in problems 9 and 12); and at the same time, they take alternative assessments seriously when it is appropriate and plausible to do that (officers who disagreed with the given assessment in problem 10 and officers who agreed with the given assessment in problem 2). Trained subjects appear to use whatever strategies are appropriate to achieve accurate assessments.

Training had no statistically reliable effects on officers' sensitivity to new information. We take this as a positive sign, indicating that training did not hypersensitize officers to ambiguous information, nor did it make them blind to the implications of that information.

Training also had no impact on confidence. This result also can be interpreted positively. The training encouraged officers to evaluate more evidence more deeply. It was possible that the wealth and quality of arguments would overwhelm rather than inform the decision-makers. This might have lessened their confidence. However, training did not have this effect. Trained officers applied new strategies to evaluate the scenarios more deeply, yet their confidence in their evaluations did not diminish.

The potential effect of several sources of individual difference were explored with several of the performance measures used in prior analyses: the number and quality of arguments officers generated, their agreement ratings, and their scores on accuracy. The principal finding was that experience in situation assessment positions correlated with greater accuracy in judgments. Training did not influence this relationship.

Subjective Evaluation of the Training

Participants who received the training were asked to comment both quantitatively and qualitatively on the instruction, during a debriefing session.

We asked participants to provide a single score representing their assessment of the training. The scale ranged from 1, denoting strongly negative, to 5, strongly positive. As shown by Figure 15, officers tended to be positive in their ratings of the course. The modal rating was 4. There was no significant difference in ratings as a function of participants' military experience when trained participants were segregated into groups at the median 10 years of military experience (Pearson $\chi^2_4 = 1.304, p = 0.861$).

Most participants (19 out of 29) reported that the situation assessment methods introduced by the training influenced their approach to the posttest. These participants reported that the training gave them "a more systematic approach," or that it helped them to "question hidden agendas or assumptions," or to account for seemingly anomalous events. Two who claimed not to have used the techniques claimed that "the training only formalized [their] thought process[es]." Four others said that training time was too short, thus, they did not learn the process well enough to apply it.

Twenty-six of the 29 participants asserted that the training might be useful in the field. Typical of these statements were the following: "It will make me more critical of my decision-making process during staff planning;" and, "I have always been the type of person who makes rash/rush decisions. This is good training for me."

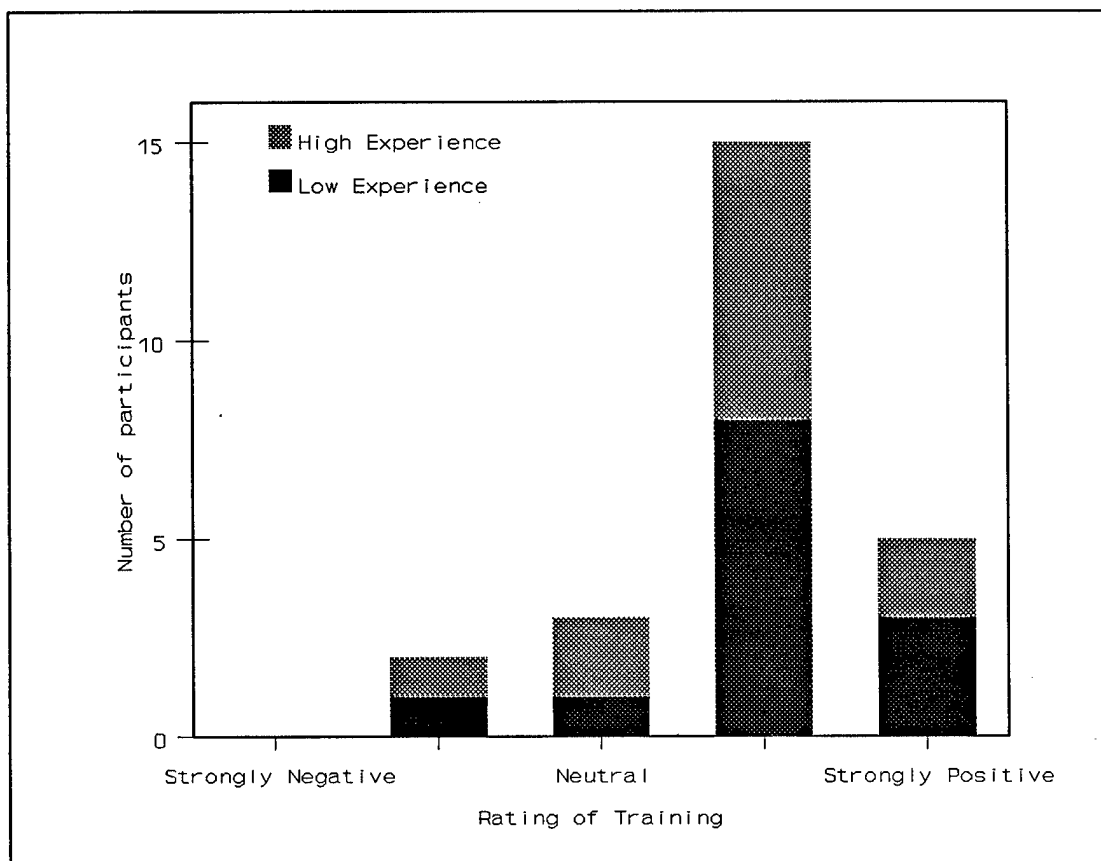


Figure 15. The most frequent participants' ratings of the training were overwhelmingly positive.

Participants also noted the importance of simple methods of critical appraisal, such as the crystal ball technique: "I will most likely use the hidden assumption method because it is quick and easy." In contrast, several participants said that the presentation of handling unexpected events was too complex, making it difficult to apply.

Only two participants stated that they would not apply the trained methods in the field. One cited the dominance of experience over training, arguing that "By this point in our careers, the experience factor drives one's faith in his assessment." The other argued that some of the specific steps trained would not survive field application, though he claimed "the questioning technique is one I have used for a number of years." One other participant was noncommittal with respect to using the methods in the field.

Participants' general recommendations concerning the training addressed two issues. Several wanted to see the course lengthened considerably. Two noted that exercises and examples could be better crafted. We agree with both critiques and note that longer training would make more detailed exercises feasible.

Finally, participants addressed themselves to the level at which the course should be taught. A few stated that the Officers' Basic Course (OBC) is the most appropriate level for the material, because it is at this level that officers form habits of reasoning about military topics. Most participants, however, felt that the Officer's Advanced Course is an appropriate forum.

There, they said, the training might be incorporated into the curriculum concerning situation assessment, course of action development, or intelligence preparation of the battlefield (IPB). Most also felt that the instruction should be extended into the Combined Arms Services Staff School (CAS3). Several suggested it also be covered in CSGC, though they cautioned that it might have limited impact late in officers' training when new reasoning techniques cannot easily displace older, habitual practices.

Conclusions

This study was designed to test the influence of training in critical thinking skills on battlefield situation assessment. We trained officers in several critical thinking strategies: how to discover unreliable assumptions by reinterpreting key events, how to notice and explain conflicting data, how to assess the plausibility of an assessment and when to develop new ones. The test itself focused on officers' ability to evaluate a specified assessment of the situation.

Training in metacognitive skills helped officers generate better accuracy of arguments without decreasing the quantity of arguments they conceived. Indeed, not only were the arguments of higher quality, but there was a trend toward greater quantity as well. Training significantly increased the quantity (as well as the quality) of arguments for officers who scored high on the pretest.

Improvements in the quality of arguments were probably related to changes caused by training in argument relevance. Training increased the proportion of arguments officers made that disconfirmed or supported a given assessment, and it decreased the number of neutral or irrelevant "arguments." Disconfirming and supporting statements give an officer leverage in evaluating assessments. Neutral statements do not. Thus, training may help officers to persuade themselves and their colleagues to retain good assessments and discard flawed ones.

Not surprisingly, therefore, training had a significant impact on the conclusions that officers drew about the situation. In particular, it improved the accuracy of their assessments, relative to the numerical evaluations of a subject matter expert (a retired U.S. Army LTG). The effect of training on accuracy was most significant on the problem for which the SME made strong predictions concerning the benefits of training.

The effects of training on accuracy were further illuminated by an examination of possible decision biases. There was some evidence for a disconfirmation bias, that is, a tendency to overreact to information that conflicts with a current hypothesis and to abandon the hypothesis too readily. Training appeared to counter this tendency, leading to more accurate assessments. In other problems, however, training appeared to counter a tendency toward a confirmation bias, that is, explaining away evidence that conflicts with a current hypothesis. In these problems, training appeared to encourage a disconfirmation bias; yet here also, the result of training was greater accuracy. The true effect of training, we think, was not to encourage or discourage these so-called biases, but to inculcate appropriate thinking strategies and the knowledge of when to use them. These strategies enable officers to generate explanations of conflicting evidence, but also to evaluate the plausibility of such explanations; at the same time, they enable officers to generate alternative assessments, and to evaluate the plausibility of such assessments. The result is that trained officers appear to confirm assessments when it is plausible and appropriate to explain away conflicting evidence; and they appear to disconfirm assessments when it is more plausible and appropriate to alter the current assessment.

We conclude that training enhanced officers' performance in situation assessment. At the very least, the ability to generate better arguments for and against a hypothesis should give officers deeper insight into the assumptions underlying assessments. This, in turn, may provide opportunities for information-gathering and shaping the battlefield that help ensure military success. These findings are encouraging, particularly given that training was very brief (90 minutes), and that control participants had spent more years in the Army than trained participants.

We were pleased to find that training did *not* influence other aspects of officers' performance. First, training did not hypersensitize officers to new information, nor did it cause them to ignore new evidence. Shifts in agreement with assessments due to new evidence were overall similar in size between trained officers and controls. Second, training did not diminish the confidence of officers in their evaluations of given assessments. This was a reassuring finding. Training introduced new thinking techniques and encouraged the participants to consider a multiplicity of interpretations of events. This might have overwhelmed them, hampered their ability to make decisions, and thus lowered their confidence in their decisions. Instead, trained officers were as confident as controls, except when controls were more confident than their level of accuracy warranted.

There remain other significant reasons to continue this research and to refine the training. Some aspects of the R/M model were not implemented in the training. One of these was the Quick Test, the gating function that determines whether time is sufficient to critique and correct an assessment, whether the stakes warrant the action, and whether the situation is novel enough that a known response may be inadequate. In a pilot study, conducted at Ft. Drum, NY, in the winter of 1993, we asked officers to rate the influence of time, stakes and familiarity on their allocation of attention to problems presented in a scenario. The results supported the importance of all three components. The Quick Test is one aspect of the model, and of the training, that warrants research in this direction.

Another aspect of the R/M model that was not introduced in training was the use of story models and other knowledge structures to organize information, for example, about enemy intent. Associated with this is the process of critiquing such models for completeness. One reason for this omission is the multitude of structuring techniques or templates already available in IPB and the Commander's Estimate. Story models and other structures may provide a useful higher-level organization for these existing representations. For example, doctrinal templates, terrain and situation templates, and event and decision templates, may fill the slots in a story structure about enemy intent, corresponding to *higher-level goals, capabilities, opportunity, intent, and actions*. The story structure might illuminate how existing templates bear on a conclusion about where and when the enemy intends to attack. Training this aspect of the model would have consumed more than the 90 minutes we had available for training in the present study, but is a worthwhile topic for future training studies.

Subsequent studies should also examine the longevity of the training effects. Our participants asserted that they planned to use the method in the field, and we hope that this is the case. However, there is the possibility that the methods trained here will be obscured just when they are needed by fading memory or the fog of war.

We are encouraged by the officers' endorsements of the training. They reported positive impressions of the course, were able to apply the lessons on the posttest, and felt it would be useful in the field. There were numerous suggestions that the training be integrated into the Officer's Advanced Course and the CGSC.

The participants identified one weakness in the course that we hope to address in the future: The training was too brief. Expanding the training would not only enable us to include the Quick Test and story structures; it would provide time for more exercises of greater complexity. These might motivate officers and help them internalize the techniques. In the unit about finding hidden assumptions, the crystal ball technique was readily learned and applied, but officers might benefit from explicit instruction and intensive practice testing the hidden

assumptions they uncover. Exercises in the next version of the training should require officers to generate intelligence gathering plans, courses of action, and compensatory plans that might prove the assumptions they reveal to be true or false. In the unit on handling the unexpected, many students said they needed more time to assimilate the somewhat complex processes involved in explaining conflicting data, evaluating the explanations, generating alternative hypotheses if the explanations are inadequate, and then explaining the data that conflict with the new hypothesis.

We conclude that training based on the Recognition/Metacognition model holds considerable promise for boosting the situation assessment skills of battlefield officers.

References

- Cohen, M.S. (1993). Three paradigms for viewing decision biases. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsombok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex.
- Cohen, M.S., Adelman, L., Tolcott, M.A., Bresnick, T.A., & Marvin, F.F. (1994). *A cognitive framework for battlefield commander's situation assessment* (ARI Technical Report 1002). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A283 631)
- Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. NY: Irvington.
- Freeman, J.T., & Cohen, M.S. (in preparation). *Methods for training cognitive skills in battlefield situation assessment* (ARI Research Product). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Gavelek, J.R., & Raphael, T.E. (1985). Metacognition, instruction, and the role of questioning activities. In D.L. Forrest-Pressley, Et al. (Eds.), *Metacognition cognition and human performance*. NY: Academic Press.
- Gordon, C.J., & Braun, C. (1985). Metacognitive processes: Reading and writing narrative discourse. In D.L. Forrest-Pressley, et al. (Eds.), *Metacognition cognition and human performance*. NY: Academic Press.
- Keeney, R., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. NY: Wiley.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. NY: Academic Press.
- Lajoie, S. P. (1986). *Individual differences in spatial ability: A computerized tutor for orthographic projection tasks*. Unpublished dissertation submitted to the School of Education, Stanford University.
- Larkin, J., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Pedhazur, E.J., & Schmelkin, L.P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

- Pennington, N., & Hastie, R. (1993). A theory of explanation-based decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsombok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex.
- Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (vol. 1). Hillsdale, NJ: Erlbaum.

Appendix A: Abbreviated Arisle Scenario

The following pages are excerpts from the Arisle scenario used as background for the pre and posttest problems (see Appendix B). The full scenario included not only the goals specified by the President and a chronology of events, but also:

- specific objectives of the operation commander
- an intelligence estimate concerning the enemy's capabilities and situation
- a report of the status of own forces
- a description of the geography, infrastructure, history, economy, and politics of Arisle.
- a detailed map of Arisle
- a large-scale map of Arisle and surrounding islands

Mission description. OPERATION POST HASTE

Just prior to the 0930 17 June meeting of the command staff for Operation Post Haste, CinCVerdiCom had passed the following Execution Order from the CJCS to the Commander, Operation Post Haste, Vice Admiral Coaler:

"The President of the United States directs that you proceed with all reasonable haste to retake the island of Arisle from the Mainlandia forces now in control of the island. It is vital to the interest of the United States and its allies that the freedom of Arisle be restored before the government of Mainlandia can gain sufficient international backing to make the restoration of full independence improbable.

You are authorized to use all reasonable force to restore Arisle with the following exceptions:

You may not enter the territorial waters of Mainlandia in any way nor fire upon any vessel, aircraft or other object that is within Mainlandia's territorial limits. You may not enter the territorial waters surrounding the Westernia island of Ebon nor fire upon any object within those territorial waters.

You may not use any destructive nuclear, chemical, or biological device in this operation under any circumstances without my direct approval.

You are to take all reasonable precaution to preclude the loss of non-combatant civilian lives. It is imperative that your government not be accused before the international community of placing other interests before those of humanitarian concern.

You have the full backing of your government and the people of the United States in this operation. God speed."

Vice Admiral Coaler then said that the CJCS had also ordered that the island be under US forces control by 2400 18 JUN. The CJCS had further stated that it is unlikely that any

significant additional combat elements could be brought to bear within this timeframe--we will have to work with the forces available.

During the discussion that followed, Commander, Operation POST HASTE decided that H-Hour would be 030018JUN. Nothing but intelligence and special operations would be conducted on Arisle prior to that time. His intent is that all reasonable effort be taken to secure the hostages prior to the outbreak of open hostilities.

The road to war. Arisle, a tropical island of approximately 345 square kilometers in the Verdi Sea, was a possession of Mainlandia from the 12th to the 18th Century when it was captured by the French during the Napoleonic Era. It remained a French territorial possession until 1947, when a peaceful revolution gained its independence. In 1964, Arisle joined the Confederation of the Liberte Islands, a primarily trade and economic assistance organization which recognizes the autonomy of its six member islands.

In 1988, a US-based corporation, Terrestria, obtained a mining permit on Arisle for the rare earth mineral, Oregonium. Only in the volcanic soil of Arisle is a deposit of the unstable metal found in sufficient concentration and a pure enough form to make mining feasible. In 1989, a new use for Oregonium was found in the manufacture of computer chips which raised the price of the already valuable metal tenfold.

In 1990, Mainlandia laid a claim before the International Court for the return of Arisle to their control, claiming it as an integral part of Mainlandia and accusing the US and Japan (which has a large pineapple plantation on the island) of economic exploitation of the tiny nation. The claim was disallowed, but Mainlandia continued to affirm its sovereignty over Arisle, but without any threat of forced repatriation until recently.

In May of 1993, Mainlandia began announced joint naval and army amphibious training maneuvers in international waters some 300 kilometers north of Arisle. In early June, the culmination of these maneuvers was a small scale amphibious "assault" of Ebon Island. This island is a tiny, uninhabited sand atoll some 200 kms NW of Arisle which belongs to Westernia. Westernia had agreed to the maneuvers, primarily because Mainlandia was to construct an airstrip on Ebon as part of the maneuvers at no cost to Westernia.

Because of the possible threat to its interests, the US decided to "show the flag" in the Verdi Sea area during Mainlandia's maneuvers. Some of the forces contingent to the theater combatant commander, CinCVerdiCom, were moved into the area to discourage any aggressive acts by Mainlandia toward Arisle and others during the maneuvers. These forces included a USN carrier battle group, units of the USA's 105 Air Assault Division, a Marine Expeditionary Unit with naval support, additional USAF forces on the air base on Madstritasia, and several Special Operations units.

At approximately 0930 hours local on the morning of 16 June, Mainlandia invaded Arisle, using Ebon Island as a staging base for the operation. The chain of events were as follows:

16JUN, 0930. A flight of ten Mi-8 HIP, two Mi-6 HOOK and two Mi-26 HALO helicopters accompanied by six unidentified fixed wing fighter aircraft coming in low level from the NW invade Arisle airspace. The helicopters land in two locations: at the air terminal and between the mine road and the American Compound 2 kms NW of the capital city of Beauqua.

By 1000, Mainlandia soldiers have complete control of the air terminal. To the south, some 100 Mainlandia paramilitary guerrillas, called 'Noclas' disembarked from the four HIPs that landed there and began rounding up US and other foreign nationals within the compound.

1015. First word of possible trouble reaches Terrestria headquarters in New York from the mine on Arisle.

1040. The mine on Arisle confirms that a Mainlandia invasion has occurred and that they are in the US compound but have not moved on the mine itself. Similar telecomms have also reached other islands of the Liberte group from Beauqua.

1045. First Mainlandia An-12 CUB cargo aircraft lands at the air terminal on Arisle. (Some 24 An-12 flights will come in during the daylight hours on 16 June bringing troops, supplies and weapon systems.)

1050. Terrestria headquarters notifies the State Department.

1105. State Department confirms report, notifies the White House and Defense Department. DoD alerts CinCVerdiCom.

1125. Emergency meeting convened at White House.

1140. State Department protests to Mainlandia ambassador and notifies the UN. CJCS issues initial warning order. Navy to begin moving Carrier Battle Group Delta, currently some 1100 kms to the ESE of Arisle, toward the island. The large US Air Force base on Madsritasia put on alert. Army told to begin preparing 3rd Bde (-) & 4th Bde (-), 105 Air Assault Division, currently undergoing maneuvers on Madsritasia, for possible movement.

1145. Carrier Group Delta informs Fleet HQ that it will be at least 24 hours before they are able to bring surface ships to bear around Arisle and 17 hours before they are near enough to launch fighter aircraft.

1150. Dodian government receives telecom from Beauqua that Mainlandia troops with armored vehicles have invaded the city.

1155. CJCS issues second warning order. Air Force given OK to begin aerial reconnaissance of Arisle. Navy given approval to prepare SEAL plt (40 personnel) for possible insertion on Arisle the night of 16117 June. CinCVerdiCom given approval to move 3rd(-) and 4th(-) Bdes, 105 AASLT Div from Madsritasia to Dodian, an island of the Liberte group some 170 kms SSE of Arisle. CinCVerdiCom given authority to form a joint task force for this operation; asked to provide continuous situation updates; to begin developing COAs for possible military action; and to identify potential needs for forces beyond those currently under his OPCOM in the Verdi Sea. CinCCentCom ordered to release Ranger Bn at MidEast Airbase to CinCVerdiCom and transport them to Madsritasia Airbase. ROE remain the same for now: fire only if fired upon.

1230. All direct communications with Arisle lost. Last communications indicated that another flight of Mainlandia helicopters were over the island, but the mine facility had still not been attacked.

1330. White House, State & Defense Department begin talks with other member nations of the

Federation of Centralia Oil Producing Nations (FOCOP) including Westernia and

Easternia to see what actions are acceptable to them. About the same time talks are begun with allied nations to determine their support for military and economic actions. Japan and France both have economic interests on Arisle as well as a number of their citizens.

1430. State Department receives reply from Mainlandia ambassador: They have all US, Japanese, French, British and Canadian citizens on the island in custody; any military action taken by the allies will put these citizens in jeopardy.

1530. Analysis of first set of reconnaissance photos over Arisle (c. 1230) reveal that Mainlandia troops are concentrated around the air terminal and in the city of Beauqua. They have at least 4 BMD-1s which were moving W along the mine road at the time the photos were taken. Satellite reconnaissance reports the movement of at least three Mainlandia cargo vessels toward Arisle, the first should make port by 1700. There is a Mainlandia air screen over Arisle and Mainlandia surface naval vessels have begun to surround the island.

1700. FOCOP representatives agree to allow allied action to regain control of Arisle, but might be forced to take "strong measures" if there are allied attacks of any kind against the sovereign territory of Mainlandia.

1715. CinCVerdiCom receives guidance from the National Command Authorities (NCA) through the CJCS to tell Carrier Group Delta Commander to take action to blockade Arisle ASAP but to "fire only if hostile intent of Mainlandia forces is evident". CG Delta Cmdr says he will be close enough to launch fighter/interceptor AIC by 0400 tomorrow. He is told that the SEAL platoon will go in between 0100-0200 tomorrow.

1730. Additional reconnaissance data reveals that the mine facility is now in Mainlandia hands. Mainlandia forces now are in all three of the towns on the island--Beauqua, Nipponia, and Mar Blanche. At least 8 BMD-1s have been identified on the island plus at least four ASU-85 self-propelled armored assault guns and four BRDM ATGM launcher vehicles. Additional troops continue to arrive via transport helicopters and An-12 Cargo aircraft from Ebon Island and the Mainlandia Channel Islands. A cargo vessel is now in the port of Beauqua, unloading troops and SA- 11 air defense weapons and equipment. A second cargo vessel is within 30 minutes of docking and two others should arrive within 4 to 5 hours (the port is equipped to handle four large cargo vessels simultaneously). In addition, a tanker appears to now be headed for Arisle. Unconfirmed reports indicate that the Noclax are holding US and allied citizens in small groups scattered throughout the island.

1830. VerdiCom reports that first units of 3&4/105 are now loading C-141s, C-5s and navy cargo ships. If all the scheduled AF and Navy support holds, all elements of the 105th currently on Madsritasia should be on Dodian by 2400 tomorrow.

2000. CinCVerdiCom names Vice Adm. Coaler, 10th Fleet Cmdr, the Commander, Joint Task Force (CJTF) for the operation (i.e., Operation POST HASTE). Rear Adm. Driver, CG Delta Cdr, named joint air/sea component commander; and COL Eager, cdr 3/105, named ground forces component commander. In addition to units of the 105th, COL Eager will have under his operational command, a Marine battalion currently steaming toward Dodian in Navy transports, and a US Ranger battalion which arrived on Madsritasia about one hour ago.

2130. Reconnaissance and other intelligence assets now report between 1000 and 1500

uniformed Mainlandian troops on Arisle plus 150 to 200 Noclans. Mainlandia is moving a strong air defense screen into position on the island, consisting of at least two batteries each of SA-11s and SA-13s plus a large number of SA-14s at the unit level. The heavy AD weapons have arrived via cargo ships and are currently being deployed. An additional two cargo ships now in port are off-loading troops and field artillery weapons, 130mm towed field guns and 122mm multiple rocket launchers (12-rd), quantity not yet determined. In addition to the 8 BMD-1s identified earlier, we can now identify 10 ASU-85 assault guns and 9 BRDM ATGM launcher vehicles which arrived via air transport plus numerous cargo and POL trucks off-loading at the port facility. Four engineer ditching machines have been spotted leaving the port facility. At least one other cargo ship is within three hours of Arisle and the tanker should dock within two hours. There are five other cargo vessels that have recently left Channel Island ports headed in the direction of Arisle. Air operations from the Channel Islands and the mainland seem to have ceased for the night. About 2100, six MIG-17 FRESCO fighter-interceptors and six HIND-D helicopters landed on Arisle. Still unconfirmed are numerous reports of locations where Noclans are holding foreign nationals, these include critical facilities such as the docks, the port tank farm and the island's electrical generating plant.

2200. Mainlandia reports willingness to release at least some of its foreign national hostages if all foreign troops are withdrawn to no less than 1200 kms from Arisle by 2400 on the 18th. The US told the Mainlandia ambassador that we are "taking it under advisement". The ambassador said that any hostile action by the US or its allies will surely result in loss of life among the hostages.

2400. CinCVerdiCom makes BG Avriel, cdr of Madsritasia Air Base, the new air component commander (Rear Admiral Driver remains as sea component commander); COL Stealth, his SOF advisor, the unified special operations component commander; and MG (You), cdr of the 105 AASLT Division, the new ground component commander for Operation POST HASTE. At the time of the announcement, you were flying from the US aboard an AWACS aircraft one hour out of the air base on Madsritasia. Arrangements were made at that time for a meeting between the Commander, Joint Task Force (Vice Admiral Coaler) and his component commanders and other principals involved at JTF HQ on Dodian at 0930 tomorrow.

17JUN, 0130. Reconnaissance and other intelligence sources now estimate about 2000 uniformed Mainlandia troops on Arisle. The ground combat force appears to be at least two battalions of the Mainlandia Army's 1st Paratrooper Regiment, under the command of BG Esau Schattu. Other troops are estimated to consist of other elements of the 1st Paratrooper Regt (i.e. mortar, ATGM, engineer, signal & logistics) plus additional artillery, air defense and logistics units. Now estimate about 150 Noclans personnel on the island and about 100 Mainlandia citizens who had been working on the island have been placed under arms and are apparently being used as a "police force" to control the citizens of Arisle. It has been confirmed that two groups of 6-8 foreign nationals each are being held by Noclans personnel in the dock area of Beauqua. There are numerous other possible hostage sites throughout the island, none of which have been confirmed at this time. Engineer activity is occurring all along the central ridge of the island and communications intercepts indicate the establishment of air defense and artillery nodes throughout the island. At the current time, a tanker and two cargo ships are off-loading at the port of Beauqua. There has been no air activity since 2200. Five other Mainlandia cargo ships apparently headed for Arisle are from six to twelve hours from port. Within the past hour, reconnaissance flights are indicating the possible withdrawal of Mainlandia surface vessels blockading the southern and eastern approaches to Arisle.

0330. 10th Fleet reports the apparent successful insertion of all 40 members of the 604th SEAL plt in ten elements of 4 each. Their primary mission is to pinpoint the location of foreign nationals being held on the island and the location of Mainlandia air defense sites.

0400. CG Delta begins launching two squadrons of F- 14 fighter-interceptors to deny Mainlandia air and sea reinforcement of Arisle. ROE for the blockade are to destroy any aircraft or vessel that "indicates any hostile intent or does not turn away from the island upon contact". (Mainlandia ambassador had been informed of these rules of engagement some 6 hours earlier.)

0530. Reconnaissance and other intelligence sources report continued Mainlandia engineer survivability activity 2-3 kms either side of the central ridge of Arisle across the breadth of the island. Over 100 weapon sites have been or are being prepared at this time. Current estimates are between 2000-2400 uniformed Mainlandia troops on the island. They are now spread throughout the island. No change in the estimated number of Noclas. Two more foreign national hostage sites are now confirmed: one in the center of Beauqua next to what is apparently the Mainlandia operations headquarters; and the other in the center of the Arisle petroleum tank farm on the SE side of Beauqua. The four hostage sites confirmed thus far account for only 20-35 of the estimated 144 foreign nationals (including 78 US citizens) believed held by the Noclas.

The Mainlandia Navy is definitely withdrawing its blockade of Arisle. The five cargo vessels that were headed for Arisle have turned back and no additional vessels have left Mainlandia ports. There has been no resumption of Mainlandian air operations except for two MIG- 17 and two HIND-D flights from Arisle itself which have taken off in the last 30 minutes, but are remaining close to the island. Only one ship remains in the port at Beauqua; the tanker that docked there some six hours ago. It probably is carrying jet fuel as ZIL- 131 POL trucks have been ferrying fuel between the ship and the air terminal.

0600. CG Delta reports the destruction of one MIG- 17 30kms SW of Arisle and the sinking of one SOVREMENYY-Class destroyer plus damage to two other Mainlandia Navy destroyers in another action by its F-18s. Both actions were taken "after the enemy had taken actions indicating hostile intent". There were no US losses. CG Delta further reports the withdraw of Mainlandia Naval vessels from the vicinity of Arisle. The first of CG Delta's surface vessels are now within two hours of Arisle waters.

0630. A reliable source within the FOCOP urges the allies to act fast to free Arisle as Mainlandia is gaining strength within the FOCOP for economic and military support on their behalf.

Appendix B: Problem Statements

Each problem statement consists of three parts: Initial information, an assessment based on that information, and new information. Part A of each problem consisted in the presentation of the initial information and the assessment. Part B of each problem consisted in the presentation of the initial information, the assessment, and the new information.

Problem 1.

The G-2 has located what appears to be all elements of the regiment's two paratroop battalions, and one company of its BMD-1 equipped airborne battalion. A 122 mm rocket launcher from the artillery battalion has also been sighted.

ASSESSMENT: Mainlandia has been able to get to the island all desired troops: one entire paratroop regiment.

NEW INFORMATION: Intelligence sources have been unable to confirm that all units of a Mainlandia paratroop regiment are present. Our air blockade resulted in five cargo vessels that were headed for Arisle turning back. No additional cargo vessels have left Mainlandia ports and no additional aircraft have landed since the blockade was initiated.

Problem 2.

Four engineer ditching machines have been spotted moving inland from the port facility. Reconnaissance and other intelligence sources report continued Mainlandia engineer survivability activity 2-3 kms either side of the central ridge of Arisle across the breadth of the island. Over 100 weapon sites have been or are being prepared at this time. Current estimates are between 2000-2400 uniformed Mainlandia troops on the island. Additional reconnaissance data reveals that the mine facility is now in Mainlandia hands. Mainlandia forces now are in all three of the towns on the island--Beauqua, Nipponia, and Mar Blanche.

ASSESSMENT: Mainlandia forces plan to dig in and fortify the towns. They believe they can withstand any US attacks with their current forces until they achieve diplomatic success.

NEW INFORMATION: There has been no resumption of Mainlandian air operations except for two MIG-17 and two HIND-D flights from Arisle itself which remained close to the island. Only one ship remains in the port at Beauqua; the tanker that docked there some six hours ago. It probably is carrying jet fuel as ZIL- 131 POL trucks have been ferrying fuel between the ship and the air terminal. The Mainlandian navy appears to be marshaling all of its warships near the Channel Islands. In addition, fighter and attack aircraft from the northern Mainlandia air bases have been concentrating at air bases along the Verdi Sea.

Problem 3.

The President of the U.S. has directed our forces to use all reasonable haste to retake the island of Arisle before Mainlandia can gain sufficient international backing to make full restoration of

full Arisle independence improbable. A reliable source within the FOCOP urges the allies to act fast to free Arisle as Mainlandia is gaining strength within the FOCOP for economic and military support on their behalf. The State Department is urging the President to use restraint since Mainlandia reports a willingness to release at least some of its foreign national hostages if all foreign troops are withdrawn to no less than 1200 kms from Arisle by 2400 on the 18th. The Mainlandia ambassador said that any hostile action by the US or its allies will surely result in loss of life among the hostages. The commander of Operation Post Haste has ordered that the US forces be in control of the island by 2400 18 JUN. The State Department believes that their intent all along was to use the hostages to avert any large-scale counterattack until they could convince the other FOCOP nations to intervene economically , or even militarily, on their behalf. Other diplomatic actions being taken by Mainlandia make it apparent that they are attempting to get other nations to Support their claim to Arisle. The Secretary of Defense has argued that diplomatic efforts will be a slow process.

ASSESSMENT: From a military action viewpoint, a deadline of 2400 18 JUN will provide adequate time for a military response to locate and secure the freedom of the hostages and then retake the island.

NEW INFORMATION: Their well-planned diplomatic offensive is apparently meeting with more success than we thought possible. Intelligence sources within FOCOP claim that the organization will most likely take actions to support Mainlandia within the next two days. The G2 believes that the Noclas may be operating on their own, and the threat to hostages lives may be very real and very imminent.

Problem 4. Omitted.

Problem 5.

The enemy parachute regiment is Mainlandia's premier direct combat force and has considerable combat experience fighting insurgent forces within Mainlandia. It is estimated that between 2200 and 2500 uniformed Mainlandia troops are on Arisle. Mainlandia has a strong air defense screen on the island, consisting of at least two batteries each of SA- 11s and SA- 13s plus a large number of SA- 14s at the unit level. They have dug in weapon positions along the entire length of the central ridge of Arisle. All of the SA- 11 SAMs and MRLs are positioned along the ridge as well as two of the three batteries of 130mm field guns. Direct observation and fields of fire for weapons along the ridge are generally excellent across the entire island except toward the NE quadrant where the teak forest obscures about one-half the shoreline from the ridge.

There are seven beach areas on Arisle that might be used for amphibious operations. Four of these are relatively small beaches of approximately 1000 meter width; the fifth is the beach in the north in front of the village of Mar Blanche. It is about 2500 meters long but the river in that quadrant cuts through it. The sixth is in front of the American compound in the bay in the south. It is about 2000 meters long but the approaches are surrounded by land. The last is an extensive beach taking up almost the whole southwest coast of the island, some 8 kms long and, in places, over 1000 meters deep. It is a nearly flat expanse of sand and, at high tide, amphibious craft would have to beach at least 500 meters from the shore.

ASSESSMENT: US forces will take high casualties in any US amphibious or air assault invasion.

NEW INFORMATION: Reports from defectors indicate that the equipment of the Mainlandian forces is in a disreputable state of repair. They have no night vision devices. Their air defense is not very sophisticated and has an operational readiness status of not more than 50 percent. Unconfirmed reports indicate that most of their 130mm ammunition may have been aboard one of the cargo ships which turned back due to the US naval blockade.

Problem 6.

Numerous reports of locations where Noclax are holding foreign nationals, including critical facilities such as the docks, the port tank [farni] and the island's electrical generating plant. It has been confirmed that two groups of 6-8 foreign nationals each are being held by Noclax personnel in the dock area of Beauqua. Two more foreign national hostage sites are confirmed: one in the center of Beauqua next to what is apparently the Mainlandia operations headquarters; and the other in the center of the Arisle petroleum tank farm on the SE side of Beauqua. We believe we have now identified the location of all 144 foreign nationals believed held by the Noclax. They are being held in groups of six to eight at locations as indicated by the green "Xs" on the sitmap. These are all positions important to the Mainlandia retention of the island. Each group is being held in the open, usually with a small tent for shelter, by four or five Noclax and are moved small distances at erratic times throughout the day and night. The Noclax have made no attempt to hide these locations.

ASSESSMENT: The enemy's main strength is the location of the hostages. It is impossible to attack their AD or artillery positions, the airfield, port facility, or water and power sources by indirect fire without almost certain hostage casualties.

NEW INFORMATION: Mainlandia has agreed to a televised International Red Cross visit with the hostages to show the world that the hostages have not been harmed. The hostages will be brought together at the airfield in two separate groups for one hour interviews with Red Cross personnel at a time to be determined. The Army Delta Force Company with 78 troops can be inserted into Arisle to provide target designation on one hour's notice.

Problem 7.

Intelligence sources have identified the location of all 144 foreign nationals held by the Noclax. They are being held in groups of 6-8 at a variety of locations. They are being kept in the open, guarded by 4-5 Noclax, and moved frequently. The Noclax are not attempting to hide the locations.

ASSESSMENT: The Noclax are providing "leverage" to the Mainlandia forces by placing the lives of the hostages in jeopardy. They do not necessarily intend to kill the hostages, even in the event of a US attack, but rather, they will use them as a human shield.

NEW INFORMATION: All hostage sites are near viable military targets. The Noclax are not commanded by MG Schattu of the Mainlandia forces, but rather answer directly to the

Mainlandia Department of International relations. In the past, the Noclans have not been hesitant to carry out terrorist acts, often killing innocent civilians. Their recent propaganda has been calling for a violent overthrow of the Arisle government, claiming that Arisle historically belonged to an ethnic group predominant in the Noclans leadership.

Problem 8.

Arisle covers an area of 345.8 square kilometers. East-to-west the widest point on the island is 24.2 kilometers. North-to-south the widest point is 17.3 kilometers. Arisle was once almost completely covered by teak forest but now the forest is restricted to the northeast quadrant of the island where some 64 square kilometers of thick teak forest remains. The only other stands of timber on the island are small groves of one square km or less, mostly along the western coast. A large pineapple plantation dominates the northwest quadrant where about 54 square kilometers of pineapples are grown. The southwest quadrant is primarily grazing land and about one-half of the southeast quadrant of the island is under cultivation. The enemy has dug in weapon positions along the entire length of the central ridge of Arisle. All of the SA- 11 SAMs and MRLs are positioned along the ridge as well as two of the three batteries of 1 30mm field guns.

ASSESSMENT: The positioning of enemy artillery and AD weapons along the central ridge offers excellent observation out to sea and along most of the shoreline as well as most of the interior land mass of the island, making invasion forces very vulnerable to enemy fire.

NEW INFORMATION: Direct observation and fields of fire for weapons along the ridge are restricted toward the NE quadrant where the teak forest obscures about one-half the shoreline from the ridge. To the south of the ridge line, there are numerous small streams that are all fordable except where the banks are steep. Although days are quite long this time of year with BMNT around 0345L and EENT around 2245L, there are about five hours per day of total darkness. The 130mm (towed) Field Gun battalion is not likely to be equipped with night vision equipment. Unconfirmed reports indicate that most of their 130mm ammunition may have been aboard one of the cargo ships which turned back due to the US naval blockade.

Problem 9.

The G-2 believes that the Mainlandia forces have been able to bring to Arisle a mixed Air Defense battalion with nine SA-1 ls and ten SA-13s. There are also approximately 50 Stinger-like shoulder-fired AD weapons scattered around the island. All are relatively unsophisticated air defense systems, with only limited acquisition and tracking capabilities. The SA-1 ls and SA-13s can operate autonomously for target acquisition and engagement. The SA- 11s are configured in triplets with only one of the three electronically active at any time. The SA-13s are paired with only one electronically active at any time. Most of the SA-1 ls are hard wired to one SA-13 to take advantage of the longer acquisition range of the on-board radar of the SA- 11. Where this is not possible, and as a backup, there is radio signal communication. The apparent increased bandwidth of the onboard electronics of both types of SAM also means that they must be attacked one by one electronically as well. Our pilots are exceptionally familiar with these systems and well-trained in evasive actions and electronic countermeasures.

The positioning of artillery and AD weapons along the central ridge offers excellent observation out to sea and along most of the shoreline as well as most of the interior land mass of the island. There are one or two alternate prepared positions for each of these weapons currently on the ridge and reconnaissance indicates that they are making frequent moves.

ASSESSMENT: The air defense of the island is well-planned but should have little impact on our air operations. The systems used are limited in capability, and our pilots have both the equipment and training to deal with them. This should provide us with complete local air superiority throughout the operation, and we should be able to attack key targets with very limited threat to friendly aircraft.

NEW INFORMATION: On a recent reconnaissance mission, one of our sophisticated fighter escorts was shot down by enemy air defense along the north shore. Before ejecting, the pilot reported that he had been painted by an radar type which he could not identify; he had attempted evasive techniques, but they were unsuccessful. A second fighter has also been reported missing after the pilot lost communications with air traffic control.

Problem 10.

The two battalions of paratroopers are spread around the perimeter of the island, apparently patrolling and defending the shoreline. It appears that one battalion's sector is south of the ridge and the other north. The southern battalion appears to have assigned sectors for all three of its paratroop companies across the breadth of the sector, each supported by 1-2 BMDs and ASU-85s and two 120mm mortars. There are five BRDM ATGM Launchers in dug-in, hull defilade positions with goods fields of fire in the south. The northern paratroop battalion apparently has one company spread out on the west and one on the east side of its sector, again supported by BMDs, ASU-85s, mortars and BRDM Launchers.

ASSESSMENT: The small enemy combat force is stretched over the circumference of the island and it should be relatively easy to inhibit reinforcement of any section of the island through friendly air power and indirect fire.

NEW INFORMATION: There are about 45 kms of two-lane primary asphalt roads outside of population centers on Arisle. These connect the three population centers and the Oregonium mine in the southwest. There are good networks of two lane secondary roads, mostly of crushed volcanic rock, in all areas of the island except the southwest quadrant. Bridges on the primary roads are generally rated at 40t or less. Bridges on the secondary roads are of much poorer quality with none rated at better than 1St. The soil has good drainage characteristics and cross-country trafficability is generally good at lower elevations except where the land is under cultivation.

Although not confirmed, it now appears that the enemy's third paratroop company of the northern battalion is with the six HIND-Ds in the woods at the base of the central peak in the NW quadrant. If this is so, it is most likely that this is an airmobile, quick reaction force for the entire regiment.

Problem 11.

Enemy communication is believed to be by FM radio, supported by three relay stations along the ridge, and by commercial landline installed 25 years ago on above ground telephone poles.

ASSESSMENT: Enemy command and control appears vulnerable. Hostage rescue operations would be enhanced significantly by destroying these relatively unsophisticated enemy C2 means.

NEW INFORMATION: A HUMINT source indicates that a modernization program has been underway in Arisle for some time to install underground telephone lines which connect all towns and industrial facilities, including the Oregonium mine and the pineapple plantation.

Problem 12.

Enemy air support on the island consists of a flight of six Mi-24 HIND-D attack helicopters and a flight of five MiG-17 FRESCO fighter/attack A/C. CG Delta reports the withdraw of Mainlandia Naval vessels from the vicinity of Arisle.

ASSESSMENT: There is no enemy naval support now available and air support for the enemy forces on the island is inadequate to the task.

NEW INFORMATION: Mainlandia still holds Ebon Island, a tiny, uninhabited sand atoll some 200 kms NW of Arisle which belongs to Westernia. Mainlandia is completing an airstrip on Ebon as part of the "maneuvers" there. The Mainlandian navy appears to be marshaling all of its warships near the Channel Islands. In addition, fighter and attack aircraft from the northern Mainlandia air bases have been concentrating at airbases along the Verdi Sea.

Problem 13. Omitted.

Problem 14.

The government of Arisle is a representative democracy with a governor and a five- member board of representatives elected by the people. The governor for the past nine years has been Quiton Pailou, who has created many reforms in education, taxation, and personal freedoms. He is greatly admired by the majority of the people, especially because the economy and standard of living has improved considerably during his administration. There is a generally cordial relationship between Pailou's government and the American Terrestria Corporation as well as the Japanese Pineapple Company.

ASSESSMENT: The great majority of the population does not support a Mainlandia take over. They can be expected to generally support a US invasion force.

NEW INFORMATION: Arisle was a possession of Mainlandia from the 12th to the 18th Century when it was captured by the French during the Napoleonic Era. It remained a French territorial possession until 1947, when it gained its independence. Recent rallies of the radical Arisle Revolutionary Front (ARF) political party have brought out large crowds with their message of "Arisle First!" A suspicious fire at the Japanese pineapple plantation last month is rumored to be the work of the ARF. Since the invasion by Mainlandia, no acts of defiance by the civilian population have been reported.